Primary HIV-1 Infection in Zurich: Molecular Studies of Transmission Biology and Epidemiology

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Philip Rieder

von

Bretzwil, BL

Promotionskomitee

Prof. Dr. med. Huldrych F. Günthard (Vorsitz)

Prof. Dr. Alexandra Trkola

Prof. Dr. Andreas Wagner

Zürich, 2011

Accomplished at the

Division of Infectious Diseases and Hospital Epidemiology

University Hospital Zürich

University of Zürich

Switzerland

Table of contents

Zusammenfassung

Mit der antiretroviralen Therapie (ART) konnte ein deutlicher Erfolg bei der Behandlung von HIV-infizierten Patienten erzielt werden. Bis jetzt ist es jedoch unmöglich mit antiretroviralen Medikamenten HIV-1 Patienten zu heilen. In den westlichen Ländern gab es in den letzten Jahren eine erneute Zunahme von HIV-1 Diagnosen, besonders unter homosexuellen Männern. In den Entwicklungsländern ist es unwahrscheinlich, dass durch ART die Epidemie von HIV-1 unter Kontrolle gebracht werden kann. Daher ist die Erforschung eines sicheren und wirksamen HIV-Impfstoffes von entscheidender Bedeutung um die Epidemie zu stoppen.
Trotz grosser Bemühungen sind empirische Ansätze zur Entwicklung von HIV-1 Impfstoffen bisher fehlgeschlagen. Zur grössten Hürde in der Impfstoffentwicklung gehört die enorme Variabilität der HI-Viren. Bei der HIV-1 Ansteckung werden jedoch nur eine oder wenige virale Varianten auf den Empfänger übertragen. Dieser genetische Flaschenhals könnte eine Achillesferse des HI-Virus darstellen. Beim Menschen wird der biologische Übertragungsmechanismus jedoch nur bruchstückhaft verstanden. Vertieftes Wissen über den Mechanismus des genetischen Flaschenhalses, sowie über die Eigenschaften der übertragenen Viren könnten wesentliche Schritte zu einem effektiven Impfstoff sein. Die virale Diversität ist einerseits ein Hindernis in der Impfstoffentwicklung, andererseits ermöglicht sie auch phylogentische Studien des viralen Genoms. Die phylogenetische Analyse der viralen Sequenzen während der primären Phase der HIV-1 Infektion ermöglicht die frühen Wirt-Virus Interaktionen zu studieren, welches sonst nur im Tiermodell möglich ist.

In Kapitel 1 wird die Komplexität der übertragenen HIV-1-Populationen untersucht und nach Risikofaktoren geforscht, welche die Transmission von mehreren Viren ermöglicht. Unsere Ergebnisse legen nahe, dass die Übertragung von mehreren genetischen Varianten nicht ausschliesslich von den Eigenschaften der Schleimhaut abhängig ist, wie dies in kürzlich publizierten Studien propagiert wurde. Die Übertragung von Virenstämmen, die den Co-Rezeptor CXCR4 benützen, ist ein seltenes Ereignis. Vorsicht ist geboten bei der Vorhersage der Co-Rezeptor-Nutzung durch genotypische Algorithmen, denn die Konkordanz dieser bioinformatischen Werkzeuge ist begrenzt.

Im Jahr 2009 traten weltweit schätzungsweise 2,6 Millionen neue Fälle von HIV-1 Infektion auf. Neuere Daten deuten auf eine hohe HIV-1-Übertragungsrate von Personen mit primärer HIV-1 Infektion. Kapitel 2 beschreibt die Transmissionsdynamik während der akuten Infektion über den Zeitraum der ART und in der chronischen Phase nach dem Absetzen der ART. Innerhalb phylogenetischer Cluster suchen wir mit Hilfe longitudinaler klinischer Daten nach Übertragungsmustern. Eine große Zahl von Neuinfektionen stammt von Patienten, welche die ART gestoppt hatten. Dies weist darauf hin, dass aktuelle präventive Bemühungen unzureichend sind. Diese Befunde sprechen für eine frühzeitige, kontinuierliche ART um die Ausbreitung von HIV-1, insbesondere bei Personen mit häufiger sexueller Aktivität, zu reduzieren.

Der Erfolg der ART in einem einzelnen Patienten wird durch die Messung viraler RNA im Blut überwacht. Aktuelle kommerziell erhältliche „Assays“ haben eine Nachweisgrenze zwischen 20 und 50 viraler RNA-Kopien pro ml Blut. In epidemiologischen Studien bezüglich des Übertragungsrisikos unter ART ist es nützlich, wenn auch einzelne replizierende Viren nachgewiesen werden können. Das HI-Virus verändert sich in hohem Masse. Daher müssen molekulare Nachweisverfahren eine breite Palette von Virenstämme erkennen. Kapitel 3 dieser Dissertation beschreibt eine Methode um phylogenetisch diverse Viren, wie HIV-1, mit einer hohen Sensivität zu detektieren.

Nach der primären HIV-Infektion nivelliert sich die Viruslast (VL) eines Patienten auf einem bestimmten Niveau ein. Dieser Wert, genannt viraler „setpoint“, kann mehr als 1000-fach zwischen Patienten variieren und ist ein Marker für den Krankheitsverlauf. Der virale „setpoint“ wird durch genetische Komponenten des Wirtes und des Virus bestimmt. In Kapitel 4 wird mit der Anwendung von „machine learning tools“ versucht, eine Signatur im viralen Genom zu identifizieren, welche mit einem tiefen viralen „setpoint“ assoziiert ist.
Unsere Ergebnisse legen nahe, dass Patienten mit der Präsenz von Viren, welchen das Aminosäure-Muster 268E/358T fehlt, eine VL von > 5000 bei Studienbeginn aufzeigen und mit geringer Wahrscheinlichkeit das Virus nach Absetzen der ART kontrollieren können. Dieses Aminosäure-Muster in HIV-1 gp120 könnte sich *in vivo* negativ auf die virale Fitness auswirken.

HIV-1-Infektionen werden von einer oder wenigen viralen Varianten oder „Quasispezies“ initiiert. Anschließend nimmt mit fortschreitender Erkrankung die virale Diversität mehr oder weniger linear zu. So ist die virale genetische Vielfalt innerhalb eines Patienten potenziell informativ für das Alter der Infektion. Genotypische Resistenz-Tests werden heute routinemäßig bei Patienten mit einer HIV-1 Infektion durchgeführt. Die dazu verwendete Methode ist eine „bulk“-Sequenzierung der *pol*-Region. In Kapitel 5 analysieren wir, ob der Anteil der mehrdeutigen Nukleotide in den „bulk“-Sequenzen eines einzelnen Patienten Aufschluss über das Alter der Infektion aufgibt. Unsere Beobachtungen zeigen, dass der Anteil der mehrdeutigen Nukleotide mit einer Rate von 0,2% pro Jahr innerhalb der ersten 8 Jahre steigt. Darüber hinaus wird gezeigt, dass ein Anteil der mehrdeutigen Nukleotide von > 0,5% starke Beweise gegen eine frische Infektion liefert (Transmission < 1 Jahr vor der Probenahme).

Zusammenfassend kann gesagt werden, dass phylogenetische Studien der viralen Sequenzen verwendet werden können, um komplexe Fragestellungen wie epidemiologische Studien und HIV-1 Übertragungsmechanismen zu erforschen. Phylogenetische Analysen können zur Identifizierung von Transmissions-Paaren verwendet werden und das Erkennen von Übertragungsmustern hilft dabei, Präventionsstrategien zu verbessern. Das genaue Verständnis der initialen Abläufe während einer HIV-1 Neuinfektion könnte entscheidend sein für die Entwicklung eines wirksamen Impfstoffes. Phylogenetische Analysen viraler Sequenzen werden zukünftig an Bedeutung gewinnen, da zurzeit neue Sequenzierungstechniken etabliert werden und die Menge an Verfügbaren Sequenzen sprunghaft Ansteigt.

Summary

Although a marked success of antiretroviral therapy (ART) in controlling human immunodeficiency virus type 1 (HIV-1) infection, ART fails to cure patients from this virus. In addition, in developed countries there is an increase in new diagnoses of HIV-1, particularly in men having sex with men (MSM). In regions where HIV-1 has hit hardest it is unlikely to eradicate the virus by ART. Therefore, a safe and effective HIV-1 vaccine is crucial to halt the epidemic. Empirical approaches to the development of HIV-1 vaccines have so far, despite huge efforts, not been successful. Among the major obstacles is the enormous genetic variability. An Achilles heel of HIV-1 may be the genetic bottleneck during HIV-1 transmission, in which only one or a few viral variants, quasispecies, from a complex viral population in the source are transmitted to the recipient. However, knowledge about HIV-1 transmission biology in humans is still insufficient. Understanding the mechanisms of the genetic bottleneck and knowledge about the characteristics of the transmitted viral strains may help to design an effective HIV-1 vaccine. Contrarily to the obstacle in vaccine design, viral diversity enables phylogenetic studies of viral sequences. The phylogenetic investigation of viral sequences obtained during the acute and recent phase of HIV-1 infection facilitate the possibility to study transmission and early host/virus interactions which is otherwise only possible in animal models.

In chapter 1 investigation of the complexity of transmitted HIV-1 populations early after infection, and assessing potential factors associated with transmission of heterogeneous viral populations has been done. Our findings suggest that transmission of complex virus founder populations may not depend solely on mucosal factors, as suggested so far by recently published work. In addition, transmission of clinically relevant CXCR4-using viral strains seems to remain a rare event and caution is warranted predicting co-receptor usage by genotypic algorithms alone, because concordance of those tools is still limited.

In 2009, an estimated 2.6 million new cases of HIV-1 infection occurred. Recent data suggests high HIV-1 transmission rates caused by individuals with primary HIV-1 infection. Chapter 2 describes the study of transmission dynamics during the acute infection, during the aviremic phase over the period of ART, and during the chronic

stage after cessation of ART. We investigated transmission patterns within phylogenetic clusters by using longitudinal clinical data. Large numbers of new infections originated from patients who stopped ART indicating that current preventive efforts are insufficient. These findings argue for early, continuous ART to reduce the spread of HIV-1, in particular in individuals with frequent sexual activity.

The success of ART in an individual patient is monitored by measuring HIV RNA copy numbers in blood plasma. Current commercial available assays have a detection limit between 20 and 50 viral RNA copy numbers per ml blood. Regarding epidemiological studies of transmission events, also under therapy, it is of interest to also detect viral RNA at single copy sensitivity. In addition HIV-1 is continuously evolving at a high rate. Therefore molecular detection assays have to detect a wide range of viral strains. Chapter 3 of this dissertation describes a method to detect phylogenetically diverse viruses such as HIV-1 with single copy sensitivity.

After primary HIV-1 infection (PHI) each individual reaches a steady state level of plasma viral load which is referred to as viral setpoint. These setpoints between individuals can vary more than 1000 fold and are surrogate markers for disease progression. HIV-1 RNA levels are influenced by genetic characteristics of both the host and the virus. In chapter 4 we aim at identifying common genetic viral signatures associated with virologic control by applying machine learning tools. Our results suggest that presence of virus lacking the amino acid pattern 268E/358T is associated with viral load (VL) >5000 at baseline of PHI and with low probability of spontaneous virologic control after cessation of ART. These residues in HIV-1 gp120 might affect *in vivo* HIV-1 fitness.

HIV-1 infection is initiated by one or a few closely related viruses. Subsequently, viral diversity increases with disease progression. Thus, the viral genetic diversity within a patient is potentially informative about the age of infection. Genotypic resistance tests are routinely done in HIV-1 infected patients by bulk sequencing of the viral *pol* region. In chapter 5, we aimed if the fraction of ambiguous nucleotides (nt) within *pol* bulk sequences obtained from an individual patient is informative about the age of infection. We found that the fraction of ambiguous nts increased at a rate of 0.2% per year within the first 8 years. Additionally, we showed that a fraction of ambiguous nts

of >0.5% provides strong evidence against a recent infection event <1 year prior to sampling.

Taken together, phylogenetic studies of viral nucleotide sequences can be used to address complex aims such as epidemiological studies and HIV-1 transmission biology. Phylogenetic analysis can be used to detect transmission pairs and the knowledge about transmission dynamics can improve prevention strategies. Better understanding how HIV-1 establishes a new infection may turn out to be a prerequisite to design efficient HIV-1 vaccines. Additionally, in the future, the phylogenetic study of viral sequences will be of great importance because new sequencing techniques are currently developed, such as 454 sequencing, and the amount of available sequences will increase exponentially.

Introduction

1. Global epidemiology

At the end of 2009, it was estimated that approximately 33.3 million people worldwide were living with HIV-1 (146), the causative agent of the acquired immunodeficiency syndrome (AIDS) (11). In the same year, 2.6 million people became newly infected with HIV-1 and 2.0 million died of AIDS. More than two thirds of the people affected by HIV-1 live in Sub-Saharan Africa (22.4 million) where HIV-1 in certain areas such as Swaziland and in Lesotho has reached a prevalence of 25%. At the end of 2009, there were around 2.2 million HIV infected people living in the developed world (6.6% of the total). In the Western industrialized countries these numbers continue to rise mainly due to the life prolonging effect of ART and the new, actual rise in HIV-1 infections among men having sex with men (MSM) (9, 32, 86).

2. Epidemiology in Switzerland

As in most industrialized countries, HIV-1 probably began to spread in Switzerland in the 1970s. Initially, the disease primarily affected MSM and intravenous drug users (IVDU), while infections of heterosexuals (HSX) only started increasing significantly in the mid-1980s, in particular also fueled by sexual spread from the IVDU population in to the heterosexual population (76).

Nowadays, over 20’000 (between 19’000 and 25’000) patients are living with HIV-1 in Switzerland with an estimated prevalence of around 0.2% (7). In 2010, HIV-1 was newly diagnosed in 609 individuals. HIV-1 was predominantly acquired by MSM (~50%), followed by HSX (~45%), IVDU (~4%) and very rarely through mother to child transmission (MTCT, <1%) (8). Even though sexual HIV-1 transmission preferentially occurs among individuals with the same risk factors (e.g. MSM) the initial phase of the epidemic among heterosexuals was strongly driven by IVDU (76). However, new HIV-1 infections decreased among IVDU due to effective prevention (e.g. needle exchange programs) (98) and subsequently the spill over of IVDU to the HSX strongly declined (76). Thus, preventive measures undertaken for IVDU resulted in a strong secondary preventive effect of a reduction of transmission from IVDU to HSX (76). In contrast, among MSM, the numbers of positive test results almost

doubled between 2004 and 2008 (8). An additional layer of complexity of disease dynamics emerges from infections acquired abroad, either by sex tourism or immigration (106, 151).

3. Evolution of HIV-1

3.1 Origin of HIV

HIV exists as two types, HIV-1 and HIV-2, which both originated in non-human primates in sub-Saharan Africa (figure 1). HIV-2 represents a cross-species infection from sooty mangabeys (55). HIV-1 strains are divided into three groups (M, major; O, outlier; N, non-M non-O), each of which represents an independent zoonosis, a transmission from simian immunodeficiency virus (SIV) in chimpanzees to humans (figure 1) (37, 147). Whereas two of these groups are rare, group M has spread throughout the world and causes more than 95% of the HIV-1 pandemic.

Long before AIDS was described in 1981, HIV-1 must have been spreading through the human population. A detailed estimate of the time and place of origin of HIV-1 and the prehistory of the AIDS pandemic is provided by comparing ancestral viral sequences obtained from stored biopsies and blood samples with current HIV-1 strains (163, 165). The phylogenetic analysis of HIV-1 group M viruses gave estimates for the date of the common ancestor between 1902 and 1921, with 95% confidence intervals ranging no later than 1933 (163). The SIV strains most closely related to HIV-1 group M have been found in infected Chimpanzee in Cameroon upstream of Léopoldville, now Kinshasa, in the Democratic Republic of Congo (figure 2) (72). Around 1900, colonialists began to penetrate this area of the remote rainforest and cities began to expand rapidly in west-central Africa causing strong migration from rural areas to these new cities. These developments were most likely important drivers for the subsequent spread of HIV-1 (133).

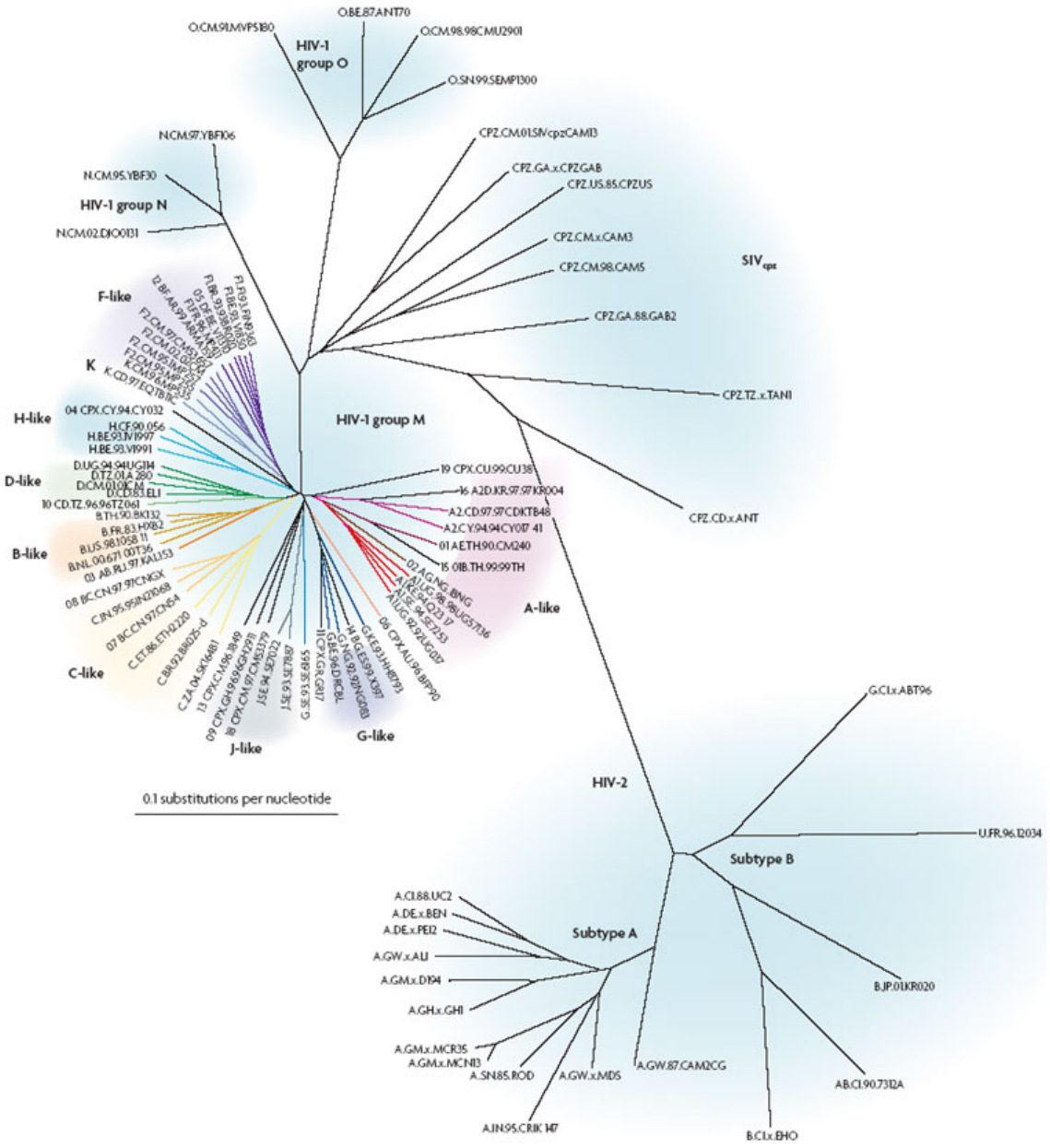

Figure 1: Phylogenetic tree of human and simian lentiviruses (adapted from (5)).

HIV-1 and HIV-2 share only 50-60% sequence identity and cluster at distinct locations on the phylogenetic tree. Whereas HIV-2 originated from a cross-species transmission from sooty mangabeys in west Africa, HIV-1 is related to SIV_{cpz}. HIV-1 is further divided into three groups, main (M), new (N) and outlier (O). The origin of group O might be related to a jump to *Homo sapiens* from *Gorilla gorilla*. HIV-1 group M can be further divided into many subtypes, which share 70-90% sequence identity, and seem to have arisen through founder events (5).

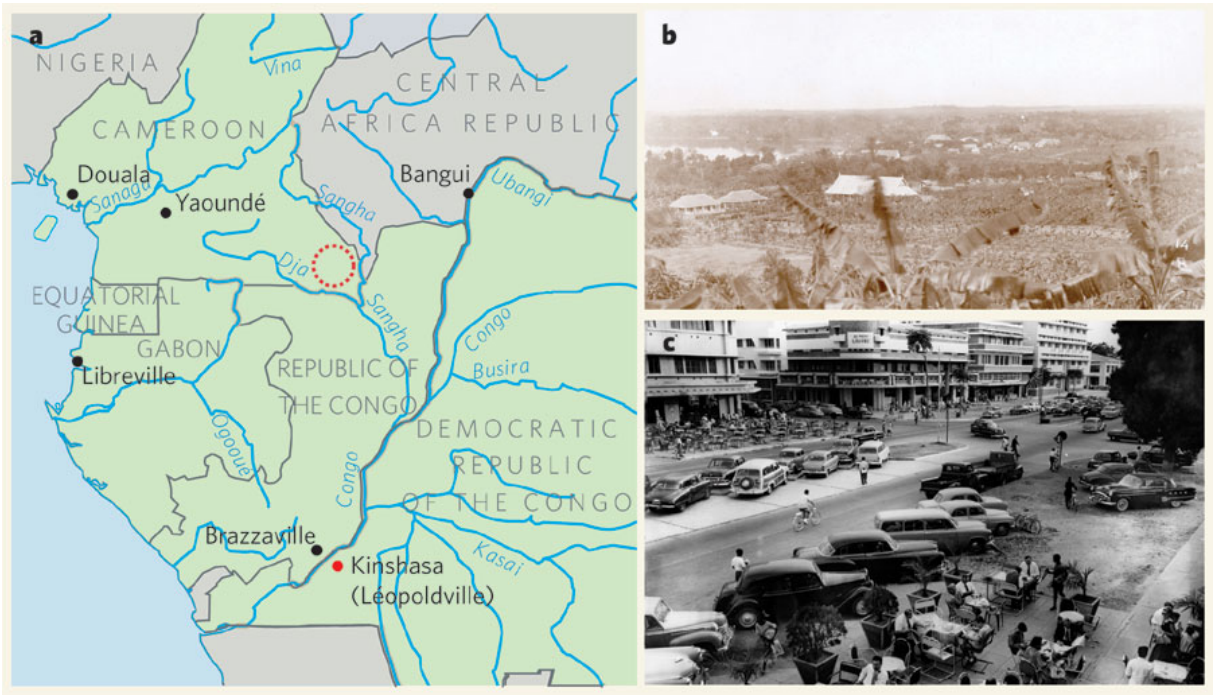

Figure 2: Origin of the HIV-1 pandemic (adapted from (133)).

a) The map of west-central Africa shows the habitat of Chimpanzees carrying the SIV strains most closely related to the viruses of HIV-1 group M (red circle) in Cameroon, major rivers, and cities with explosive population growth in the twentieth century. A possible explanation how SIV jumped to humans would be the human migration into this tropical region along the rivers, an exposure to SIV infected bush meat and a subsequent transmission and spread in the dense population of the growing town, Léopoldville. **b)** Léopoldville in 1896 and **c)**, Léopoldville around 1955 (133).

3.2 Global spread of HIV-1

The expansion of HIV-1 across Africa and throughout the world has been accompanied by one of the most rapid evolutionary rates described for a human pathogen, aside from Hepatitis C virus (43). Phylogenetic analysis identified nine subtypes within group M, including A-D, F-H, J and K, which diverge by 20-30% in the envelope (*env*) region of the genome (figure 2). In addition, a multitude of inter subtype circulating recombinant forms (CRFs) and additional unique recombinant forms have been recognized (17). Most subtypes and CRFs are present in west-central Africa. In contrast, in many countries, a specific viral subtype may be most prevalent due to founding of the local epidemic by a few individuals (figure 3) (6, 77). Nowadays Subtype C accounts for more than 50% of infections worldwide whereas subtype B accounts for just approximately 10% (figure 3) (51). The prevalence of non-B subtypes has been increasing in Western European countries, where the epidemic was initially dominated by subtype B (77). In Switzerland, the majority of HIV-1 infections still occur with subtype B viruses, but other clades such as A, C, CRF01_AE and CRF02_AG are increasingly observed, due to migration of infected patients or sex tourism (figure 4) (106, 151).

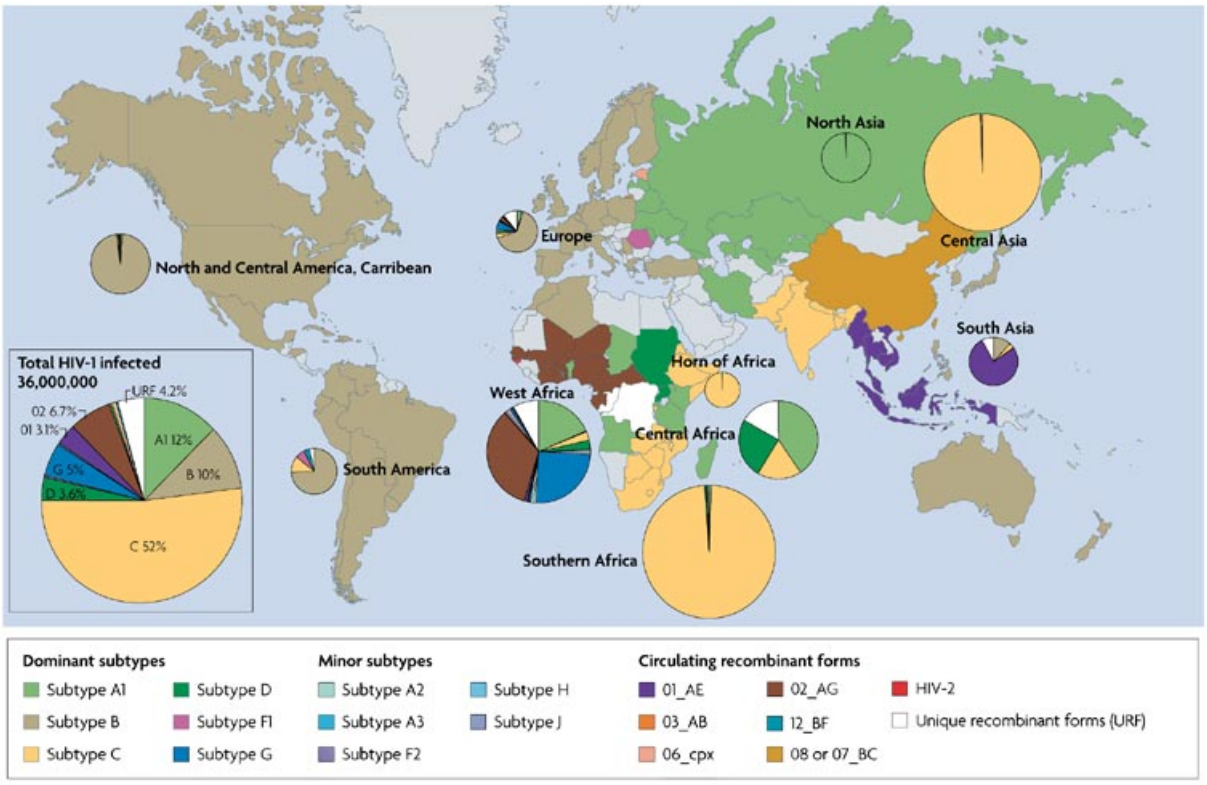

Figure 3: Geographical distribution and prevalence of HIV-1 subtypes in 2004 (adapted from (5)).

The pie charts depict the proportion of each subtype or CRF in each geographical region. The size of the circles is proportional to the HIV-1 prevalence in the particular region (5). The countries coloured grey have a low level of HIV-1 or data were not available. In 2004 subtype C accounted for 50% of the epidemic worldwide. Subtypes A, B, D and G contributed for 12%, 10%, 3% and 6%, respectively. The CRF_AE and CRF02_AG each were accounted for 5% of all infections. All CRF together were responsible for 18% of cases.

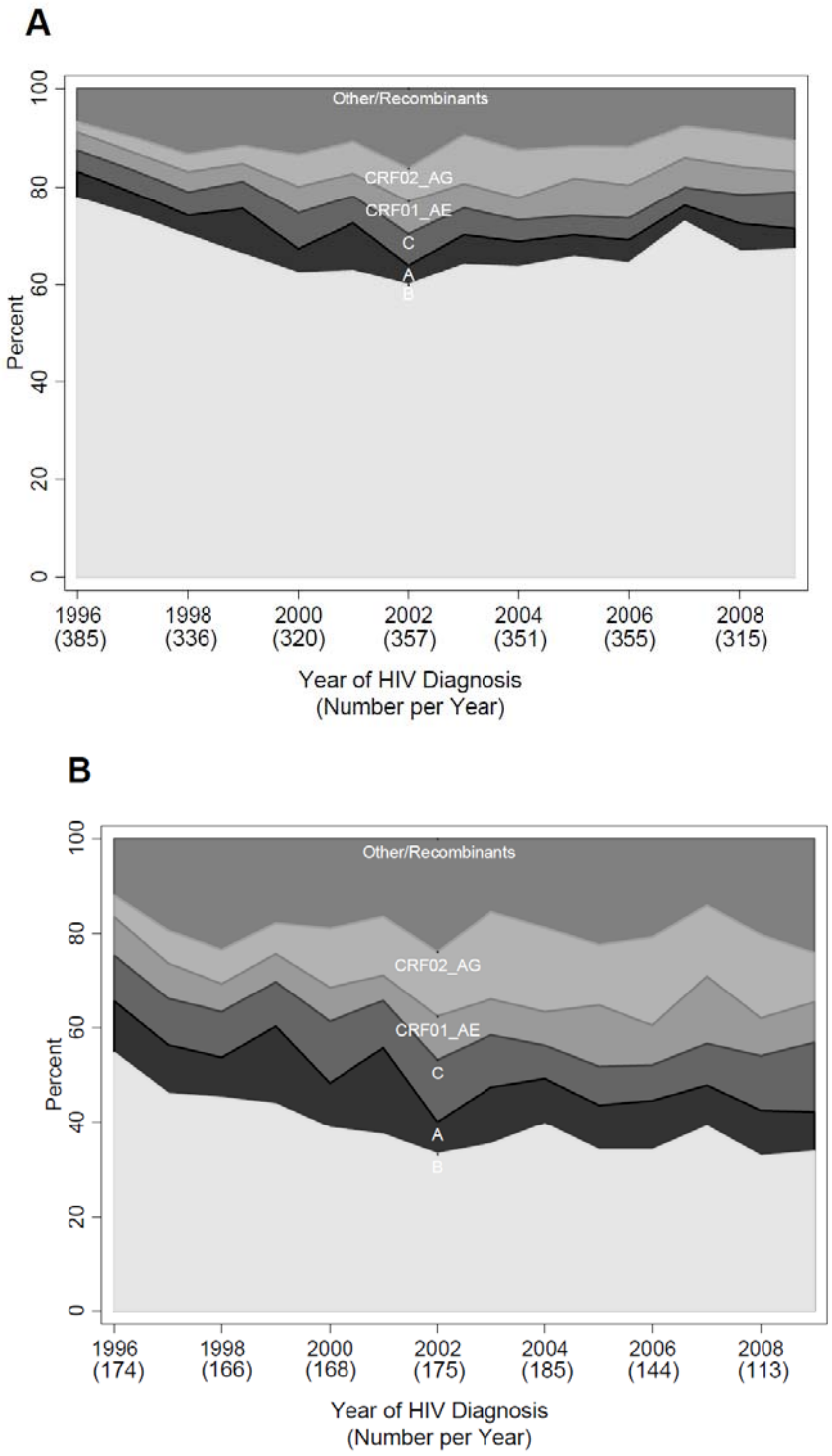

Figure 4: Time trends in distribution of subtypes in the Swiss HIV Cohort study (adapted from SHCS).

Panel A: All modes of HIV transmission (n=4767); **Panel B:** Heterosexual contacts only (n=2197). Graphs for intravenous drug users (low numbers) and homosexual men (no discernible time trend) are not shown (151).

3.3 Natural course of HIV-1 infection

The natural course of the HIV-1 infection is divided into three clinical phases: the acute phase, the chronic phase and AIDS (figure 5) (reviewed in (79)). The acute phase starts with an infection and lasts 4 to 8 weeks. During this initial stage, around 40-90% of patients show symptoms of the acute retroviral syndrome (ARS) (3). These large varieties of unspecific symptoms appear approximately 1-3 weeks following exposure to HIV-1 but can vary significantly between subjects and last from one to several weeks (127). During acute HIV-1 infection, there is often a marked decrease in $CD4^+$ T-cell count and extensive viral replication. Subsequently, concurrent with the appearance of HIV-1 specific $CD8^+$ T-cells plasma viral load falls abruptly (74). $CD8^+$ T-cells may contribute to the control of HIV-1 infection through both innate and adaptive mechanisms (130, 162). The drop in viral load may be driven by other factors as well, such as loss in activated $CD4^+$ T-cell targets or other immune-mediated mechanisms (61). Neutralizing antibodies develop during acute infection and thus may contribute to viral clearance (36, 136). Nevertheless, both quantity and quality of the humoral immune response improves during the course of infection and broader neutralization activity is predominantly found in chronic infection (95, 96). Taken together, the immune response is gradually building up, leading to a decrease of viral replication and to equilibrium between the immune response and viral replication after around 3 months. This viral setpoint, following resolution of the acute infection, is a strong predictor of long-term disease progression rates and characteristic for the second phase of HIV-1 infection (90). The viral setpoint is defined by host (~20%) and viral (~50%) characteristics (4). During this stage of infection the patient stays asymptomatic for a time span of around 8-10 years in 70-80% of the cases (intermediate progressors) and for a few years or over 12 years in rapid or late progressors, respectively (reviewed in (79)). During the chronic stage $CD4^+$ lymphocytes gradually decline over time and drop below a threshold of approximately 200 $CD4^+$ cells/μl. Thereby the patient enters the third phase of infection and becomes vulnerable to AIDS-defining illnesses such as opportunistic infections (caused by bacteria, mycobacteria, parasites, viruses and fungus) and certain tumors.

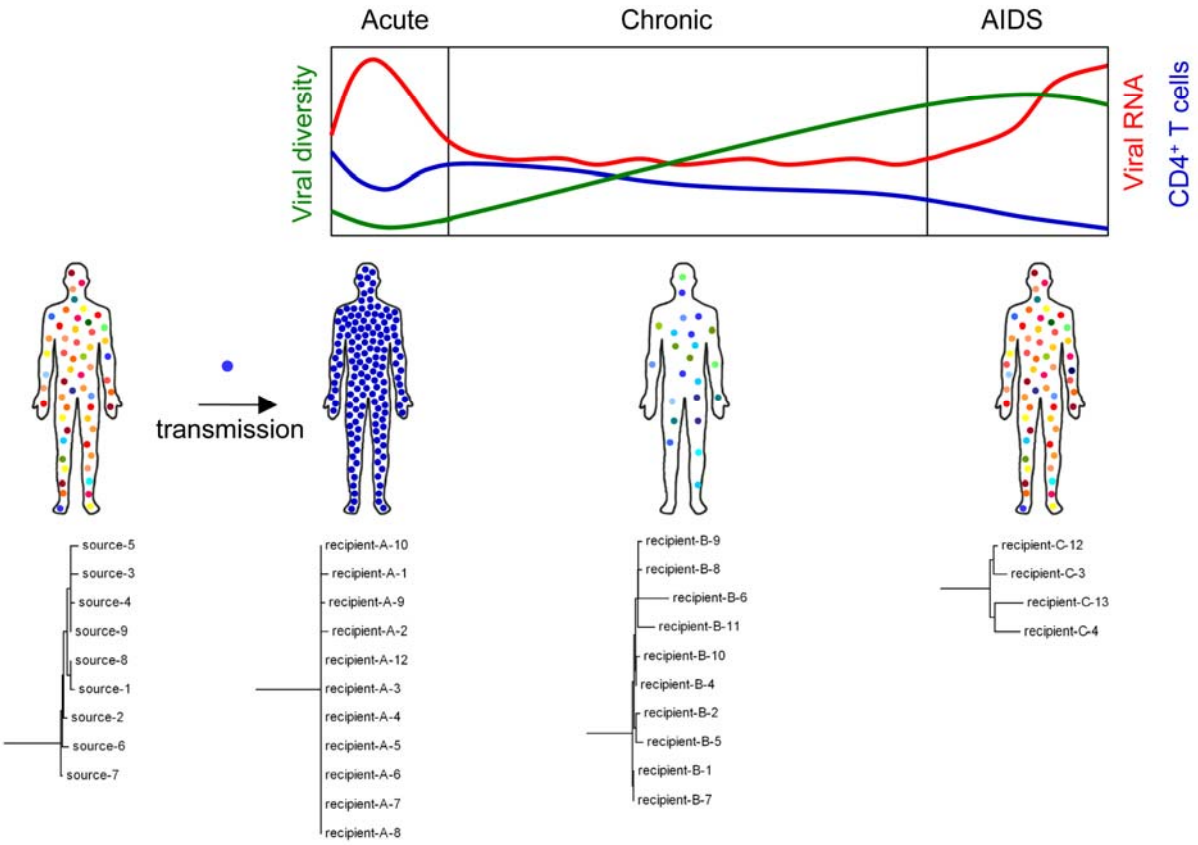

Figure 5: Natural course and viral diversity during HIV-1 infection

In the acute phase, HIV replicates rapidly to maximal viremia (red) accompanied with a decline in blood CD4+ T cell count (blue). Later, blood CD4+ T cells rise again to a moderately subnormal level because evaded T cells move back from lymphoid tissue into the blood and proliferation is increased. In the chronic phase, viral load rises and CD4+ T cell levels decrease slowly. When CD4+ T cell levels drop under a certain limit (200 cells/µl blood) opportunistic infections occur, defining AIDS.

An individual is typically infected by one or a few HIV-1 clones (depicted by the small blue circle), which dramatically increase in copy number but not in genetic diversity during the first 1-2 months. Following the acute phase genetic diversity (green) increases at a relatively linear rate with the length of infection until a plateau is reached in the AIDS phase. Viral diversity is also indicated as colored circles within the human contour and in phylogenetic trees of full length *env* sequences.

3.4 Intra-patient HIV-1 evolution

An individual is typically infected by a few HIV-1 strains (described in more detail in section 4.2 and figure 5) (71). During disease progression, this founder population rapidly evolves into a mixture of different viral strains, referred to as a quasipecies. Viral diversity builds up mostly in a linear fashion, until a plateau is reached (75, 132). During AIDS, even a decrease in viral diversity has been observed (figure 5) (132).
The extreme heterogeneity during later stages of HIV-1 infection is a result of rapid viral turnover, a high virus burden in an infected individual, recombination and the error-prone nature of the reverse transcriptase (RT) enzyme which lacks proofreading activity (56, 113, 114, 157). Besides stochastic forces (genetic drift) natural (positive) selection is the driving force of intra-host evolution. HIV successively fixes mutations that allow the virus to evade immune responses and to acquire optimale fitness in a given host. Host immune selection pressure are generated mainly by HIV-specific neutralizing antibodies, T-helper cells (119) and by cytotoxic T lymphocytes (CTLs) (14).
Genetic viral diversity between HIV-1 infected patients varies substantially. It remains unclear whether HIV-1-related disease progression occurs more rapidly in patients harboring a viral population with low or high diversity. Various studies have described that patients with low viral genetic diversity have a delayed disease progression (68, 85, 87). Others argued that higher viral diversity may induce a broader immune response which subsequently could control the infection (160). The loss of $CD4^+$ T-cells and development of AIDS would be primarily due to the increasing antigenic diversity that, beyond a threshold, exceeds the capacity of the immune system to control the virus (160). In this model, the antigenic diversity is the cause of AIDS.
It is also a matter of debate whether the complexity of the founder population affects viral replication in the host and subsequently the viral setpoint (123). This steady-state level of viral replication is generally established during the recent phase.

4. HIV-1 transmission

4.1 Transmission mode of HIV-1

Transmission of HIV-1 can occur via unprotected sexual intercourse with an infected partner, injection or transfusion of contaminated blood or blood products, sharing unsterilized injection equipment that has been previously used by someone who is infected and maternofetal transmission (during pregnancy, at birth, and through breastfeeding) (reviewed in (59)).

4.2 Transmission biology of HIV-1

The SIV rhesus macaque non-human primate (NHP) model of vaginal transmission of HIV-1 provides a window through which one can observe early infection prior to systemic infection (93) in a way that is not possible in HIV-1 infection. From the NHP model we have the following picture of the earliest stage of infection (figure 6) (reviewed in (46)): Virus can cross the mucosal barrier by different mechanisms preferentially at sites of single layered epithelium such as the endocervix and the transformation zone in the female genital tract. These mechanisms for HIV-1 transmission across mucosal epithelium may include direct infection of epithelial cells, transcytosis through specialized microfold (M) cells, epithelial transmigration of infected donor cells, uptake by intraepithelial Langerhans cells and circumvention of the epithelial barrier through physical breaches (reviewed in (134)). Initially infected cells are relatively small numbers of “resting” $CD4^+$ T-cells that lack detectable markers of activation. This founder population then undergoes a local expansion during the first week of infection to generate sufficient virus and infected cells which afterwards disseminate via lymphatic drainage to establish a self-propagating infection in the genital draining lymph nodes. Subsequently, the virus spreads systemically by dissemination into the bloodstream and infection of the secondary lymphoid organs.

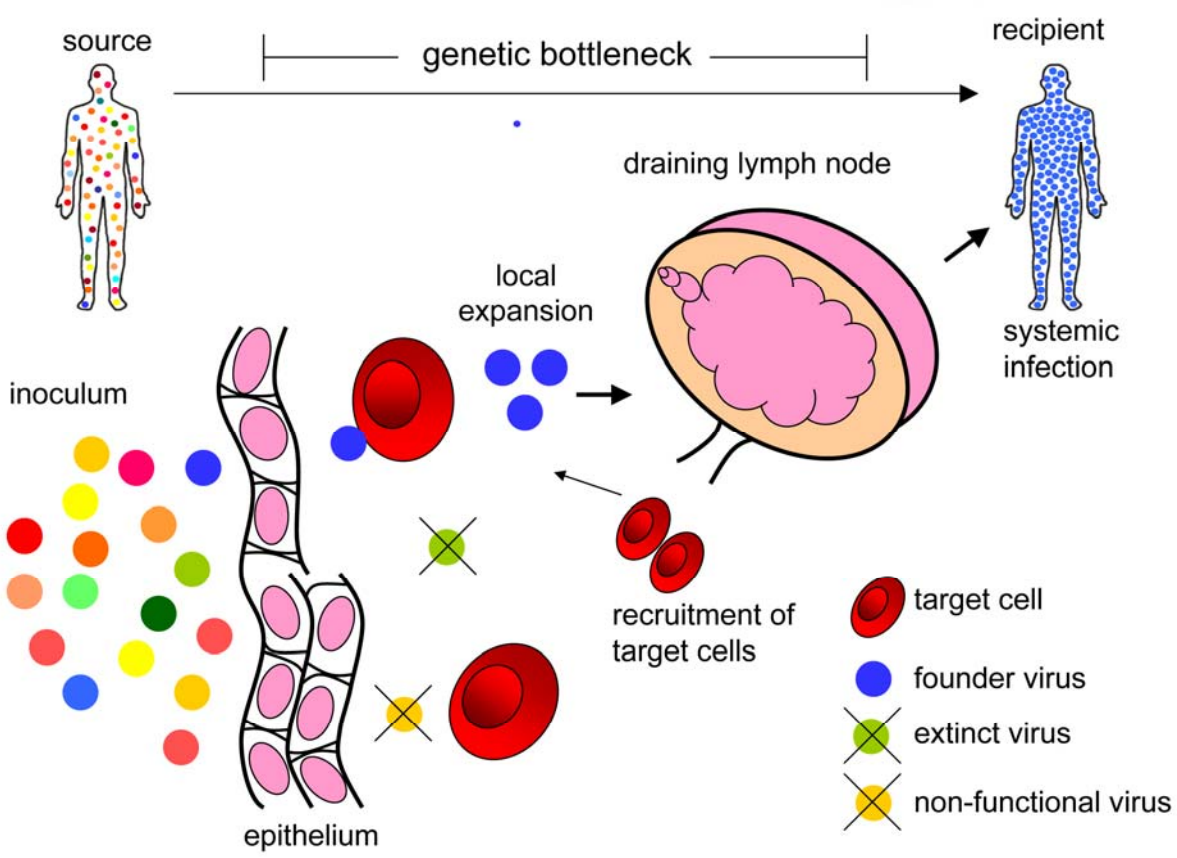

Figure 6: Early events during vaginal HIV-1 transmission

Virus crosses the mucosal epithelial barrier to establish a small founder population. Other transmitted viruses extinct because of the low target cell availability or are non-functional viruses. The small founder population expands locally using the influx of new target cells recruited through outside-in signalling. The local expansion disseminates via lymphatic drainage to genital lymph nodes where expansion quickly produces more virus and subsequently disseminate through the bloodstream to establish a self-propagating infection in secondary lymphoid organs and get systemic.

4.3 HIV-1 selection during transmission

Despite a swarm of closely related viruses present in the donor, there is a strong genetic bottleneck associated with transmission so that only a limited number of variants are transmitted to the recipient (figure 5 & 6) (28, 118, 159, 161). Recent studies suggested that HIV-1 transmission results from productive infection by a single transmitted/founder virus in ~80% of HIV-1 infected HSX (2, 45, 71), in ~60% of HIV-1 infected MSM (71, 83) and in ~40% of IVDU (10). The likely transmission of a single viral variant in many cases provides a potential window of opportunity for vaccines and microbicides, which, at least in a majority of cases, would need to protect against a very small incoming viral dose of limited diversity.

For this reasons it is of interest to understand factors associated with the acquisition of multiple viral variants versus single variants, not only because this may have potential implications on vaccine design but also because high diversity following transmission has been associated with more rapid disease progression (41, 124). Apart from the mode of transmission and the anatomy of encountered mucosal barrier, sexually transmitted infections (STI) were also found to be associated with increased frequency of multiple variant transmission in some studies (2, 45, 124). Together, the mucosal epithelial barrier may be responsible for the genetic bottleneck during transmission (figure 6). However, a genetic bottleneck has also been observed within IVDU, therefore strongly suggesting that additional selective mechanisms independent of the mucosal barrier most likely are relevant (10).

It is still an unresolved question whether the bottleneck during transmission is a stochastic event or whether active selection of specific variants occurs during HIV-1 transmission. Support for the later hypothesis was provided by studies showing that newly acquired variants of non-B subtypes have shorter and/or less glycosylated envelope glycoproteins than those present in the transmitting partner or among chronically infected patients (20, 28). Over the course of infection glycan restricted envelope glycoproteins develop a glycan shield due to immune pressure of neutralizing antibodies (156). In addition to this observation, newly acquired viruses are more closely related to ancestral sequences compared with the predominant variants circulating in a chronically infected host (52, 122). Similarly, although patients in a later stage of infection often harbour mixtures of viral variants that use chemokine receptor 5 (CCR5) and CXC chemokine receptor 4 (CXCR4) receptors for

host cell entry (figure 7), predominantly CCR5 tropic viral variants (R5) are transmitted (reviewed in (12)), although here studies vary significantly (1, 23, 25, 116, 166). Together, adaptation within a host may be counterproductive for transmission efficiency.

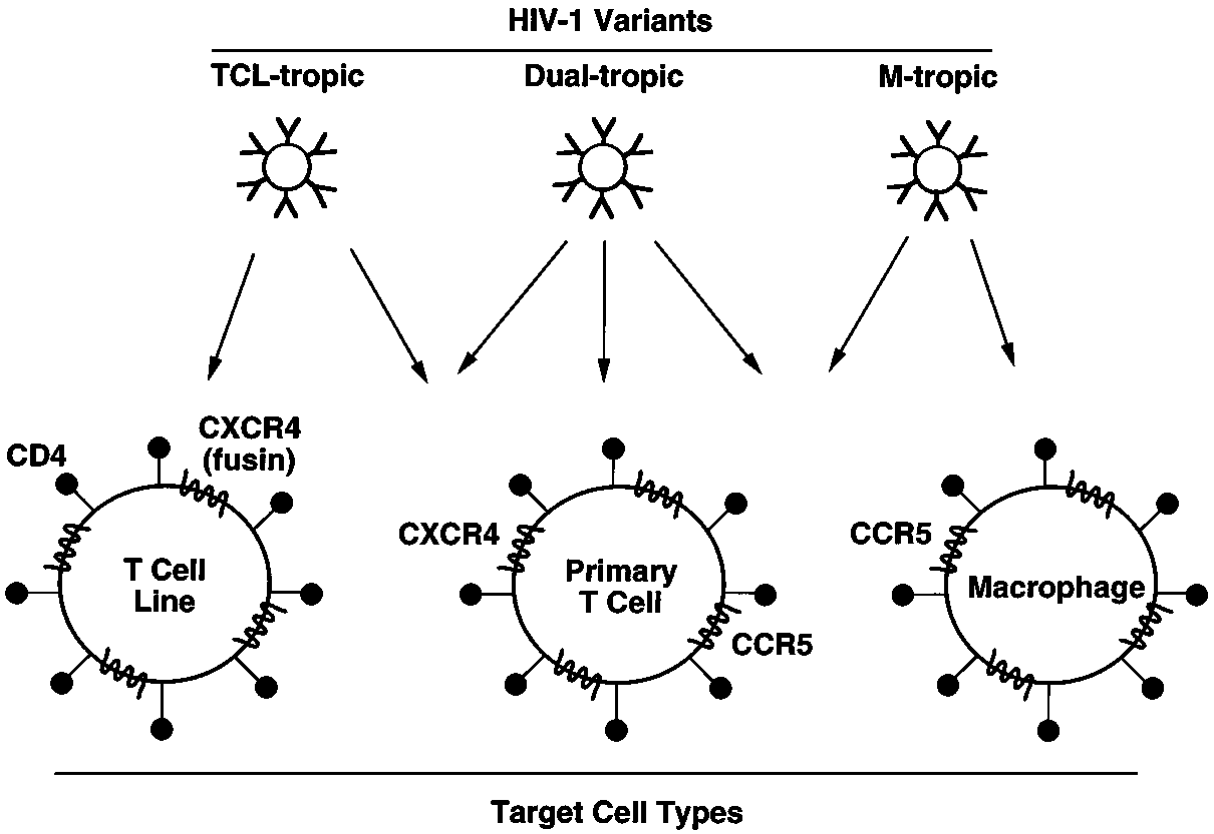

Figure 7: Model for corezeptor usage and cell tropism (adapted from (12)).
T-tropic (X4) strains are specific for CXCR4 and M-tropic (R5) viral variants use CCR5 as corezeptor. Dual-tropic strains can use both corezeptors for cell entry (12).

5. Antiretroviral therapy (ART)

5.1 HIV-1 structure

HIV-1 belongs to the genus lentivirus, a member of the family of retroviridae. Its genome comprises two positive copies of single-stranded RNA, both of which contribute to replication. The viral genome is around 10 kb in length and encodes for 15 processed proteins (figure 8 & 9). Those proteins that are derived from *gag* (group-specific antigen), *pol* (polymerase) and *env* (envelope) are structural and enzymatic factors that are common to all retroviruses. In addition, HIV-1 encodes two regulatory proteins, Tat (transcriptional transactivator) and Rev (regulator of expression of virion proteins) and four genes that encode accessory proteins, nef (negative factor), vif (virion infectivity), vpr and vpu (viral proteins R and U, respectively) (reviewed in (35)).

The viral envelope protein

Env is synthesized as a polypeptide precursor (164). During the transport through the Golgi apparatus *env* undergoes extensive glycosylation which is required for correct folding and conformational stability. In the trans-Golgi network the *env* gp160 is cleaved into gp120 and gp41 by the cellular protease furin (reviewed in (78)). Gp41 is the transmembrane unit of the viral envelope protein which is noncovalently linked to the surface unit gp120. During virion budding the gp120/gp41 complexes are incorporated into the virus envelope. Functional glycoprotein spikes on HIV-1 particles consist of trimers of these complexes. Gp120 consists of a signal peptide followed by different conserved (C1/C2/C3/C4/C5) and variable regions (V1/V2/V3/V4/V5) (141). The conserved regions are hidden from the immune response by tertiary and quaternary structure of *env*. Gp120 is further protected by the “glycan shield” obstructing antibody recognition of both variable loops and conserved sites (156). Only antibodies that react with the intact trimer are considered to bear neutralizing activity (reviewed in (62)). However, both gp120 and gp41 monomers, as also gp160 precursor proteins, are released upon disintegration of infected cells (161). Additionally, gp120 monomers can further dissociate from spikes due to the inherent instability of the trimer (164). Within gp120, the V3 loop, the

V1/V2 loop, and the CD4-binding site were identified as target sites for neutralizing antibodies (reviewed in (62)). However, V1/V2 domains protect the V3 loop on neighbouring gp120 subunits within the envelope trimer (120).

The variable regions of the *env* are the major determinant of virus coreceptor usage and cell tropism (discussed further in chapter 5.2 and figure 7). V3 amino acids in X4 strains have a significantly higher positive charge than R5 isolates (34). The presence of a positively charged amino acid at 11^{th} and/or 25^{th} positions of the V3 loop is a predictive marker for the X4 phenotype (117).

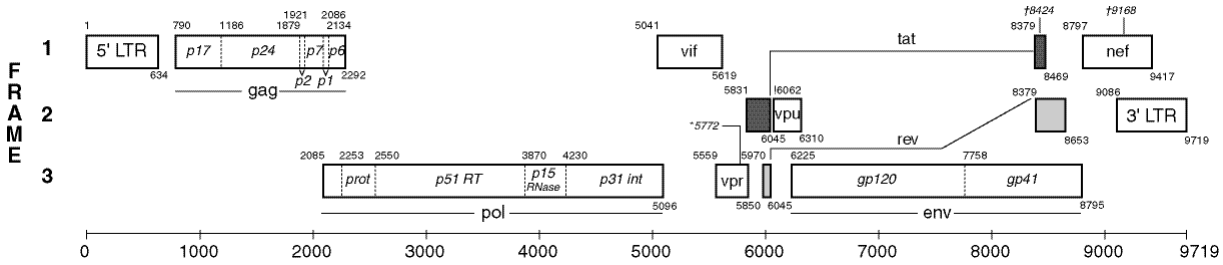

Figure 8: Genetic organization of HIV-1 (adapted from www.hiv.lanl.gov).

Open reading frames are shown as rectangles. Gag and Gag-Pol polyprotein precursors are processed by the viral protease into subunits: matrix (MA, p17), capsid (CA, p24), nucleopcapsid (NC) p1 and p6, protease (PR), reverse transcriptase (RT), which contains RNaseH and integrase (IN). Env (gp160) is cleaved by cellular proteases, such as furin, into gp120 and gp41 moieties. The tat and rev spliced exons are shown as shaded rectangles.

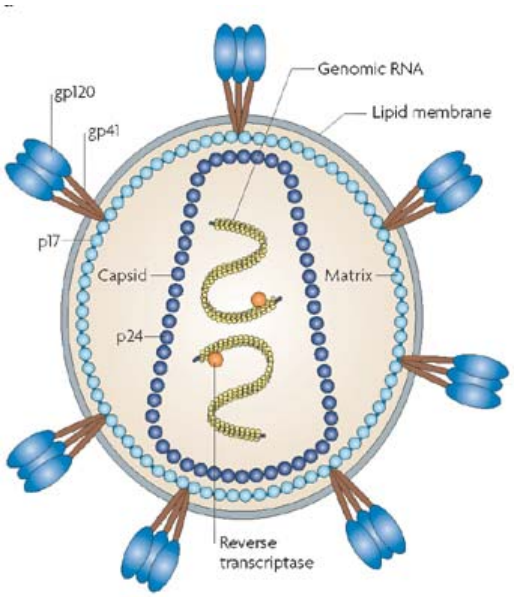

Figure 9: The structure of an HIV-1 virion (adapted from (69)).

The structural proteins form the virion which is about 100-130 nm in diameter. Its outer layer consists of a lipid membrane derived from the former host cell in which trimeric gp120-gp41 complexes are embedded. The inner surface of the membrane is lined by the matrix which interacts with the cytoplasmic tail of gp41. The capsid proteins forms a conical core. The 2 HIV-1 RNA copies are surrounded by nucleocapsid proteins. In addition, proteins necessary for reverse transcription and integration of the provirus are incorporated into the virion (69).

5.2 Viral life cycle

The initial step of the viral life cycle starts with the attachment of the virus to the host cell by binding of the viral *env* protein gp120 (SU protein) to the CD4 receptor inducing extensive conformational changes in gp120 that exposes the coreceptor binding site of the chemokine family. The major co-receptors required for entry of HIV-1 are the chemokine receptor molecules CCR5 (R5 HIV-1 isolates) and CXCR4 (X4 HIV-1 isolates), which define the tropism of a HIV-1 strain (figure 7). Binding to the coreceptor results in another structural alteration of gp120 exposing the N-terminal part of gp41 transmembrane subunit (TM protein). This part, also known as the fusion-peptide, mediates the fusion between the viral and host membranes.
After fusion of the two membranes, the viral core enters the cytoplasm and undergoes uncoating thereby the single-stranded RNA is released and converted into double stranded DNA by the viral reverse transcriptase enzyme. The newly formed DNA copies, called proviruses, are then actively transferred to the nucleus of the host cell where the viral enzyme integrase (IN) inserts the proviral DNA into an active part of the host genome. The LTRs flanking the proviral DNA contain enhancer and promoter sequences, with binding sites for several transcription factors and a polyadenylation signal. In productively infected host cells, the proviral genome is transcribed into mRNA and translated into proteins. Then, viral assembly takes place and virus particles are released from the host. After virus budding from the surface, maturation of the virus particle proceeds. Gag and Gag-Pol polyproteins are proteolytically cleaved by the protease. Cleavage of Gag results in MA (matrix), CA (capsid) and NC (nucleocapsid). Cleavage of the Gag-Pol polyproteins, which are generated by ribosomal frameshifting during translation of unspliced RNA, results in the enzymatic proteins IN, RT and PR. After maturation the virus is ready for another round of infection (reviewed in (35, 69, 97, 107).

5.3 Overview of antiretroviral drugs

So far, 26 single or combination preparations from five classes of drugs are available: nucleoside and nucleotide analogs (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), entry inhibitors and integrase inhibitors (INs) (143).

Nucleoside analogues target the HIV-1 enzyme reverse transcriptase and act as alternative substrates, thereby competing with physiological nucleotides and inducing the abortion of DNA synthesis, as phosphodiester bridges can no longer be built to stabilize the double strand. AZT and 4dT are thymidine analogs, while FTC and 3TC are cytidine analogs; ddI is an inosine analog, which is converted to dideoxyadenosine; and abacavir is a guanosine analog. There is a high degree of cross-resistance between nucleoside analogues (54). A classical ART regime has always nucleoside/nucleotide analogues as a “backbone”. Today the most important ones are TDF+3TC and 3TC+ABC in developed countries (reviewed in (59)).

As with NRTIs, the target enzyme of NNRTIs is RT, however, they bind directly and non-competitively to the enzyme. The resulting complex blocks the catalyst activated binding site of RT. Not only is the risk of resistance high, it can also develop very rapidly. One point mutation (K103N) of the binding site is enough for resistance to the entire class of first generation NNRTIs. Despite the potential problems with resistance, they are extremely effective when combined with NRTIs (reviewed in (59)).

PIs fit exactly into the active enzyme site of the HIV-1 protease thereby preventing proteolytic cleavage of the gag-pol polyprotein, which results in non-infectious virus particles. There is a considerable degree of cross-resistance between protease inhibitors (22). Ritonavir is a very potent inhibitor of the isoenzyme 3A4, a subunit of the cytochrome P450 hepatic enzyme system. Inhibition of these gastrointestinal and hepatic enzymes allows the most important pharmacokinetic parameters of almost all PIs to be significantly increased, or “boosted” (73).

There are 2 entry inhibitors available so far. T-20 (Enfuvirtide) inhibits fusion of virus and cell (reviewed in (99)). Maraviroc (MVC) is a chemokine (C-C motif) receptor 5 (CCR5) antagonist and selectively inhibits the replication of CCR-5 tropic (R5) HIV-1 variants via an allosteric mechanism after binding to a transmembrane CCR5 coreceptor pocket (19). CCR5 antagonists do not display activity against chemokine (CXC motif) receptor 4 (CXCR4)-using HIV-1 variants. As a consequence, the presence of detectable X4 or R5/X4 dual-tropic viruses, either as dominant strains or as mixtures of CCR5 and CXCR4-using strains, has been associated with treatment failure (30). Evaluation of HIV-1 tropism is therefore necessarily before treatment initiation with MVC.

5.4 History of ART

In retrospect, the history of ART can be divided into four phases. The first phase started in 1987, when the NRTI AZT became available. Other nucleoside analogs followed a few years later. Administered as mono or dual therapy, those drugs did not seem to provide durable efficacy and had a modest impact on disease progression and mortality of HIV-1 infected patients (80).

The first phase ended in 1996, when the drug class of PIs was introduced (48). At around the same time, it became apparent that a “cocktail” of several drugs simultaneously yielded a much better and more durable response (figure 10) (47). Treatment guidelines recommended ART as a combination of three drugs from two classes (2 NRTI and 1 PI). Subsequently, morbidity and mortality declined enormously among patients on combination ART (cART) (100).

With increasing knowledge of side effects of ART many treatment recommendations were revised. Instead of “hit hard and early”, it changed to “hit HIV hard, but only when necessary” (49).

Nowadays, the pattern of ART initiation is changing again towards earlier treatment. 2010 recommendations of the international AIDS society - USA panel suggest ART for asymptomatic patients with CD4 cell count ≤500/µl, for all symptomatic patients, and those with specific conditions and comorbidities. ART should be considered for asymptomatic patients with CD4 cell count >500/ µl (140). However, the question of “when to start?” is still under debate and will be discussed in the next paragraph.

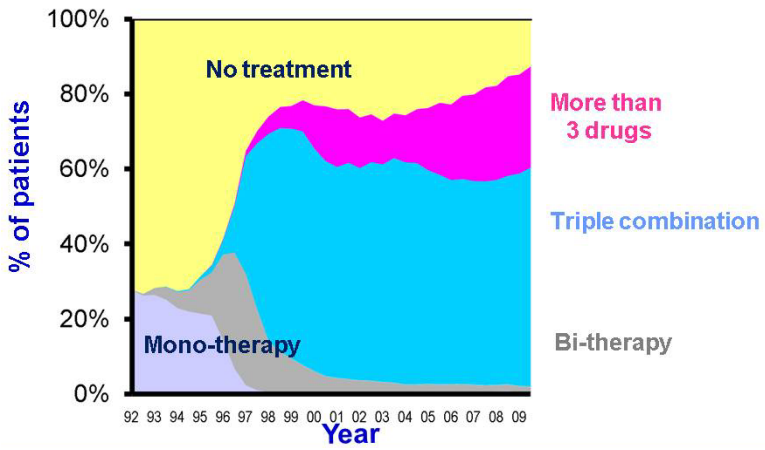

Figure 10: History of ART (adapted from SHCS).

This graph illustrates the increasing use of multiple (i.e. triple or more than 3 drugs) combinations therapies in the Swiss HIV Cohort study (SHCS) since 1996 (135).

5.5 Clinical management of acute HIV-1 infection

The optimal clinical management of primary HIV-1 infection (PHI) to date is not known. There are several potential benefits and disadvantages of initiating ART during PHI (early ART).

Early ART can have a positive effect on disease markers (40, 57, 58, 70, 101, 129), but whether this has long-lasting clinical significance remains unclear. Viral setpoint establishes within the first 6 months after infection and is prognostic of disease progression. Early ART reduced the viral setpoint after cessation of ART in some studies (39, 139). But it is not known how long the effect lasts. In the ZPHI study it looks like the initial benefit after approximately 1 year is lost after 2-3 years (unpublished data; vonWyl, Günthard).

The acute retroviral syndrome (ARS) is characterized among others by mononucleosis-like symptoms, typically lasting 2 weeks but potentially continuing for ≥ 10 weeks (127). Early ART can suppress viremia, thereby lessening direct viral effects.

$CD4^+$ T cell depletion in the gut occurs very early after infection and is a crucial event in the pathogenesis of HIV-1 infection (88, 148). The relationship between early ART and preservation of gut associated T lymphocyte depletion is under debate (50, 89, 128, 149).

There is an estimated more than 10-fold increased risk of transmission during the PHI compared to the chronic stage of HIV-1 infection (60, 154). In addition, genetic analyses linked between 5 and 50% of new infections to sequences obtained from other patients during PHI (15, 82). Therefore early ART has gained increased attention as a prevention method.

6. Factors associated with increased infectiousness and susceptibility

HIV-1-associated morbidity and mortality have declined among patients with access to ART (29). However, this success was not matched by similar reductions in HIV-1 transmission. Current data even suggest a recent increase in HIV-1 diagnoses in resource-rich settings, particularly among MSM (7, 32, 86). The reasons for this phenomenon may be intricate. The introduction of cART in 1996 has completely changed the HIV-1 epidemic (figure 10). HIV-1 infection was no longer a sentence of death but rather a manageable chronic disease. In response to the beneficial effects, sexual risk behavior and other sexually transmitted infections (STIs) increased among MSM (142). Due to “condom fatigue” and lack of availability of an effective vaccine or microbicide within a reasonable timeframe there is a need for re-examination of the approach to control this virus.

The infectiousness of HIV-1 in male genital fluid together with the susceptibility of the host, the type of sexual practice, and the viral load are most likely major determinants of sexual transmission. The risk of acquiring HIV-1 is directly correlated with the level of virus circulating in the source blood (115, 154). Highly active antiretroviral therapy (HAART) reduces viral replication and thus lowers the rate of HIV-1 transmission (41, 115). The Swiss Federal Commission for HIV/AIDS suggested that seropositive individuals, which are on ART with undetectable viral loads for more than 6 months and do not have other sexually transmitted diseases, do not transmit HIV-1 (150). Although HAART reduces viral loads in both blood and seminal compartments, HIV-1 RNA can still be detected in seminal fluids (136). Therefore it remains under debate if undetectable plasma HIV-1 RNA lowers the risk of viral transmission through seminal fluid not only on a population level but also at an individual level.

Both acute and late stages of HIV-1 infection are associated with high viral loads and, therefore, may disproportionately contribute to the spread of the epidemic (15, 108, 115). The importance of the acute infection period for subsequent transmission events is additionally related to the unawareness of the infection and potentially virus-specific properties that confer fitness for transmission.

Genotypic antiretroviral resistance testing has become commonplace, resulting in an accumulation of HIV-1 *pol* sequences, forthcoming for adequate phylogenetic reconstruction of transmission events (63). Various studies using diverse methods suggest that 5-50% of incident cases are acquired from newly infected individuals

(15, 60, 111, 154). Various definitions of acute phase, difficulties in the estimation of exact infection date and neglecting the transient nature of the acute stage may produce such varying results (16). Altogether, the success of “treatment as prevention” is highly dependent on the understanding of the transmission dynamics and necessity of identifying and characterizing of the source patient.

STIs greatly increase the efficiency of HIV-1 transmission by increasing both the infectiousness of, and the susceptibility to HIV infection. STIs can be divided into four categories: those that produce no mucosal signs or symptoms (e.g. HIV, hepatitis B); those that produce mucosal inflammation (e.g. Neisseria gonorrhoeae, Chlamydia trachomatis); those that produce genital ulcers (e.g. herpes simplex virus (HSV) 1 and 2); and those that cause epithelial changes (e.g. human papillomavirus). STIs that cause ulcers generally increase shedding of HIV in the genital tract and can also affect HIV levels in semen (reviewed in (33)). STIs that cause inflammation increase the concentration of HIV in semen (140) and cervical fluids (38). There are several possible mechanisms for the increased risk of acquiring HIV-1 that is seen in individuals with STIs. The main target cells for HIV-1 are CD4+ cells immediately below the epithelial layer (44) that are rendered accessible by breakdown in the mucosal integrity by STIs (94). Both, ulcerative and non-ulcerative diseases lead to cellular infiltration to the site of infection, increasing the potential target cells for HIV infection (24). In addition, cervical CCR5 expression is increased in woman with STIs (105).

7. Diagnosis of HIV-1

Early diagnosis during the critical stages of PHI represents a tremendous opportunity for treatment and prevention interventions. However, acutely infected patients are difficult to identify. A popular strategy to study persons with PHI is to prospectively follow up subjects at high risk for seroconversion, such as MSM, IVDU, sex workers, and HIV-discordant couples. The pitfalls of this approach include (1) reduced risk of HIV acquisition with safe sex counselling, (2) the cost of following up HIV-negative persons, and (3) the difficulty in studying subjects early after transmission. Most persons can not be seen more than once a month. Another strategy to detect PHI-patients is targeted screening of symptomatic PHI. Most individuals who acquire HIV-1 develop nonspecific signs and symptoms. However, such cohorts have the limitation of enrolling patients with a severe ARS, which have a potential for greater risk of disease progression (81). Additionally, physicians may recognize an ARS within MSM but overlook it in HEX where HIV-1 is less common.

Additionally, there is no absolute or widely accepted definition of acute and recent HIV-1 infection (16). Acquisition of HIV-1 is followed by a series of events that is characterised by different patterns of viral antigens and antibody responses and that can be used not only for diagnosis but also for staging of the infection (figure 11) (31). However, laboratory staging, such as the Fiebig system, is dependent on the sensitivity of the used assays, which has markedly improved in the last years. The definition is also affected by the variability of viral replication kinetics and the varying host immune responses among individuals. A further limitation to the Fiebig staging system is that the durations of each stage and diagnostic assays are based mainly on clade B HIV-1 infections and viral constituents, respectively (31).

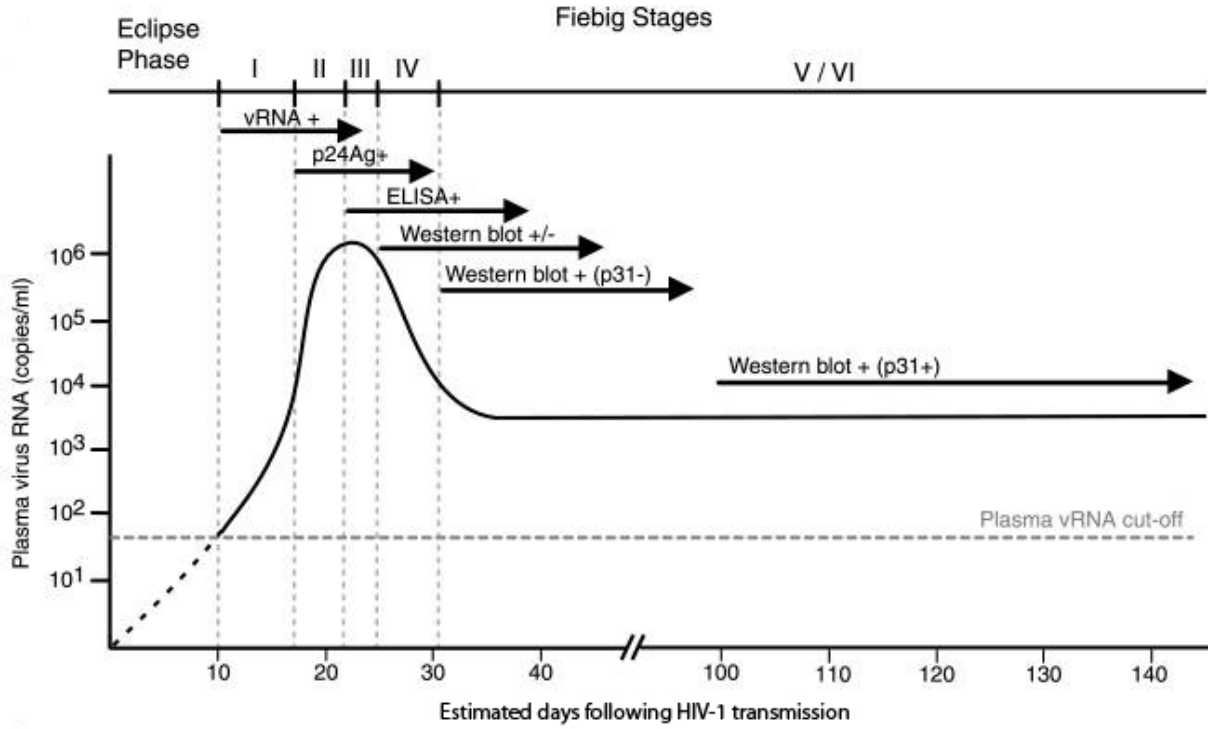

Figure 11: Laboratory staging of early HIV-1 infection (adapted from (31)).

This figure shows the temporal appearance of HIV-1-specific laboratory markers following HIV-1 infection according to the classification system of Fiebig et al. (31). The eclipse phase is defined by the interval between transmission and first detection of viral RNA in the plasma and generally lasts about 10 days, with a range of approximately 7 to 21 days. There are six stages defined based on laboratory markers: stage I (vRNA positive, p24 antigen and antibody negative), stage II (vRNA and p24 antigen positive, antibody negative), stage III (ELISA antibody positive, Western blot negative), stage IV (ELISA positive, Western blot indeterminant), stage V (ELISA and Western blot positive, p31 integrase antibody negative), and stage VI (ELISA, Western blot, and p31 integrase antibody positive). The mean durations of Fiebig stages I (7 days), II (5 days), III (3 days), IV (6 days), and V/VI (70+ days) are indicated.

Reprinted by permission from (31).

Detection of viral RNA

Viral load is defined as the level of HIV-1 RNA copies/ml of plasma. The documented association between the numbers of HIV-1 virions in the blood during the asymptomatic phase, referred to as viral setpoint, and the clinical course of infection has led to the use of molecular tests (figure 12) as the standard of care for treating patients in developed settings (91). The utility of molecular detection of HIV-1 RNA in body fluids and tissue has increased especially for monitoring patients on ART. Additionally, viral load testing has provided valuable information about the risk of transmission, and it can be used to diagnose HIV-1 prior to seroconversion. While genetic variability has a limited impact on serologic detection of HIV-1, molecular diagnostic assays have been more challenging (103). While primers are relatively forgiving in respect to target variation (21), as little as one mismatch in a fluorescent hydrolysis (FH-probes; TaqMan® -probes) probe's target sequence can greatly perturb the results of quantitative PCR by underestimating copy numbers (26).

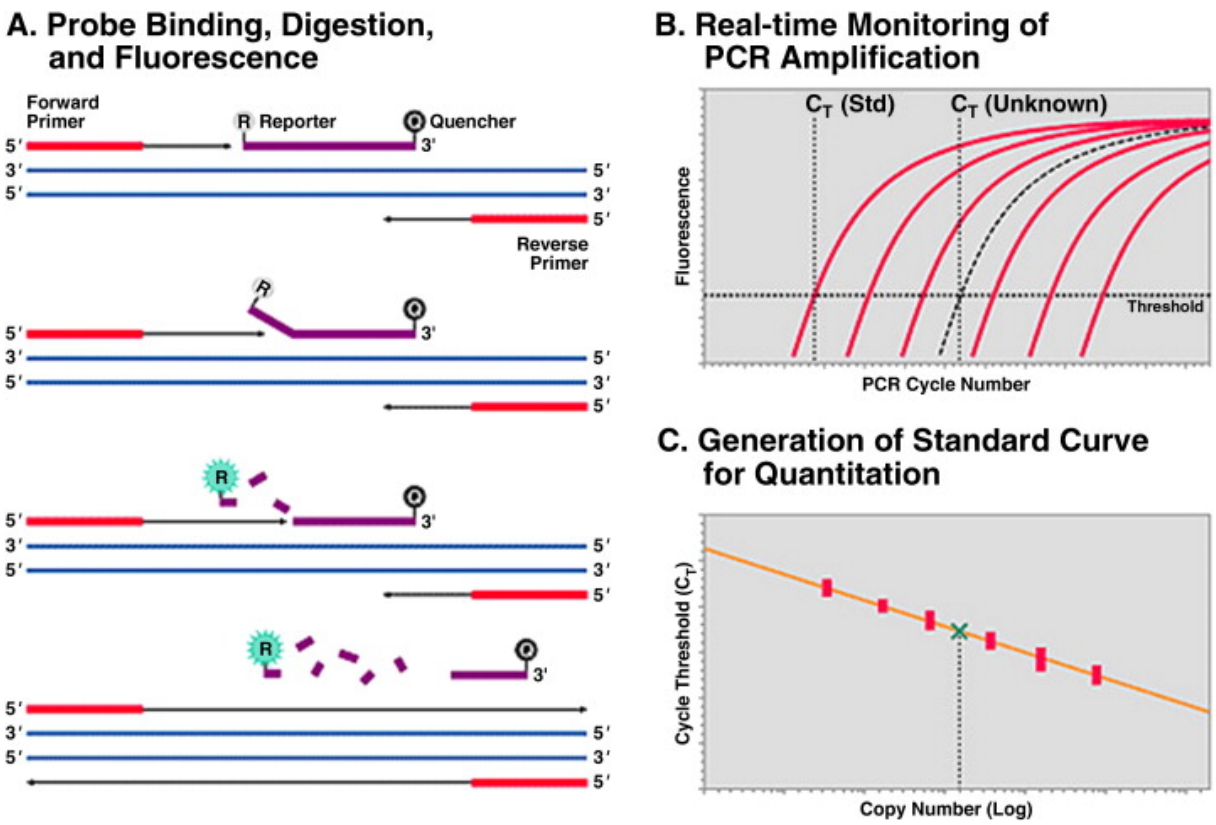

Figure 12. Schematic illustration of TaqMan® technology for viral genome quantification (adapted from(153)).

The TaqMan®-probe is a dual labeled DNA sequence of a short nucleotide stretch, with a fluorescence reporter at the 5' end and a quencher at the 3' end. Due to the proximity of them, the fluorescence emitted by the reporter is suppressed by the quencher. **A)** During amplification, a TaqMan®-probe binding to the DNA template is digested by DNA polymerase, releasing fluorescence. B) Six external standards are amplified with an unknown sample in parallel. When the fluorescence exceeds the background signal, the cycle number of PCR is recorded and used for quantification. C) The more HIV-1 RNA in a sample, the earlier the fluorescence exceeds the background and the lower the cycle threshold (C_t) values are. The viral load of the unknown sample is indicated by the green cross.

Reprinted from (153) copyright (2010) with permission from Elsevier.

8. ZPHI-study

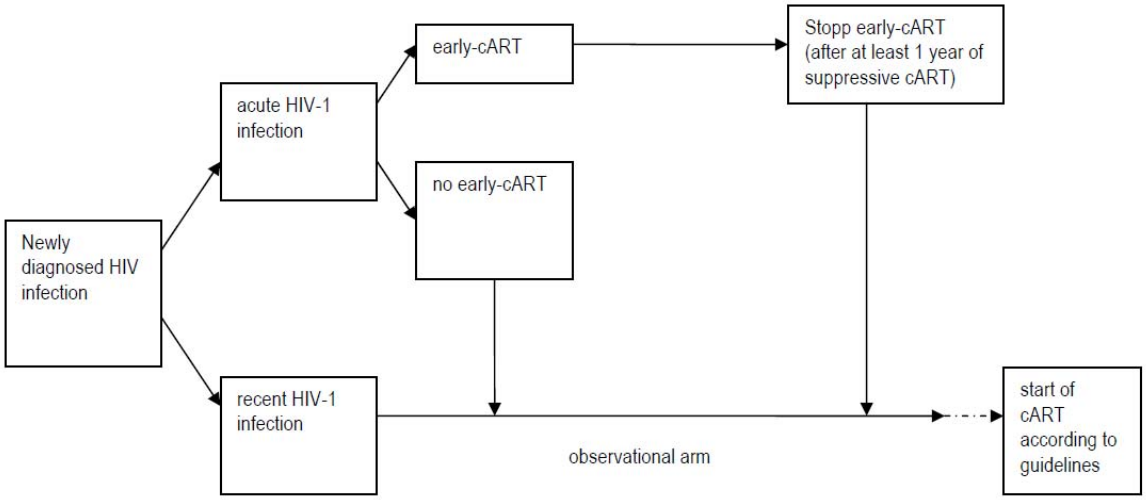

Figure 13: Schematic outline of the ZPHI-study

Newly diagnosed patients who present with a documented acute or recent HIV-infection are enrolled. Acute is defined as:

- an acute retroviral syndrome (ARS) (3) and negative or indeterminate Westernblot (Wb) in the presence of a positive p24 Ag and or detectable plasma HIV-1 RNA.
- documented seroconversion with or without symptoms within 90 days

A recent infection is definded as:

- possible ARS, positive Wb and detectable HIV-RNA, and a negative HIV-gp120 avidity (13, 117), respectively detuned assay(63)
- documented acute infection, however, referral to our center more than 90 days after estimated date of infection

The Zurich primary HIV infection (ZPHI)-study enrols and longitudinally follows patients with documented primary HIV infection (http://clinicaltrials.gov). This is an open label, non-randomized, observational monocenter study at the University Hospital Zurich, Division of Infectious Diseases and Hospital Epidemiology, which started in 2002. Patients are offered early ART if treatment starts within 90 days after HIV-1 infection. After one year of suppressed HIV-plasma viremia (<50 copies/ml) patients can choose to continue early ART or switch to the observational arm (figure 13). The primary aims of the study are to describe the epidemiology, test the effect of early ART, investigate early events of host-virus interactions, search for host and viral factors which are associated with HIV-1 transmission and influencing viral setpoint. The ZPHI-study includes around 55% of all primary HIV-1 infected MSM in the canton Zurich and is therefore representative for this transmission group.

9. SHCS-study

The Swiss HIV Cohort study (SHCS) was already established in 1988 and some data going back to 1985 have been collected retrospectively. It is an ongoing multicenter, prospective observational study for interdisciplinary HIV research in clinical, translational, basic, epidemiological, and social sciences. The data collection is strictly anonymous and written informed consent is mandatory for inclusion and also for patient genetic analysis. Today the SHCS includes HIV infected individuals aged ≥ 18 years and is estimated to cover around 51% of the cumulative number of HIV infections and 68% of people living with AIDS in Switzerland (www.shcs.ch). In the SHCS, a standardized protocol is used for data sampling. Socio-demographic and behavioural data are collected and various serological laboratory tests are routinely performed at enrolment. At each semi-annual follow-up visit laboratory, clinical data and ART is documented in detail. Moreover, in the SHCS plasma/serum samples (twice a year) and viable cells/cells pellets (once a year) are stored and available for research purposes (131).

The Swiss HIV Cohort Study drug resistance database is a recent extension of the SHCS (149). Genotypic drug resistance testing has been performed routinely in Switzerland since the year 2000 and all tests are entered into a central database (Integrated Database Network System, IDNS (SmartGene, Zug, Switzerland)).

According to the guidelines, drug resistance testing is recommended prior to initiation of ART and following a viral rebound on therapy (53). For research purposes genotypic drug resistance testing is also performed retrospectively with frozen samples. In the SHCS, one third of samples was tested at baseline, 30% have received resistance testing after initiation of ART and 27% of all patients ever enrolled in the SHCS have never been tested due to loss of follow-up or death (vonWyl, unpublished data).

10. Methods to characterize viral genomes

HIV-1 exists as an evolving population in infected individuals. There are several techniques to characterize the variants making up a viral population each with its own assets and drawbacks.

Heteroduplex tracking assay (HTA)

The migration of a discrete double-stranded DNA molecule through a non-denaturing electrophoretic gel is influenced by the degree of base-pairing of the double helix. DNA molecules with exactly complementary single strands run more quickly than identically sized DNA strands with mismatched bases. Also, an insertion or deletion in one strand causes a perturbation in the mobility of the molecule within the gel matrix. A genotype can be assigned on the basis of the separation of the homoduplexes and heteroduplexes by electrophoresis (27). This avoids the need for extensive sequencing but has the pitfall that no sequences are available for further analysis. Therefore, this technique will not be discussed further in this thesis.

Cloning and sequencing

The most common practice to obtain multiple sequences is by performing RT-PCR on a region of the viral genome, cloning the amplified products into a vector, and randomly selecting a number of clones for sequencing.
If the number of RT-PCR templates is low, it is unlikely that all sequences subsequently obtained by cloning will be derived from different input templates resulting in re-sampling of individual genomes in the population. However, a previous study in our laboratory revealed that there was no correlation between viral diversity and plasma HIV RNA (68). Thus, no evidence was found that low viral diversity was due to low input of HIV RNA. Because primer DNA sequences used in PCR are pre-defined, PCR imposes a selection which may underestimate actual intra-patient diversity (84). Contrarily, PCR-based recombination has been observed, generating additional diversity (84). Nucleotide misincorporation during generation of cDNA by error prone reverse transcriptase results in overestimation of viral diversity. In addition the amplicon will contain Taq-induced nucleotide misincorporations as it is

molecularly cloned prior to sequencing . However, the average misincorporation rate is very low (0.18%) (68).

Single genome sequencing (SGS)

Single genome sequencing (SGS) permits individual cDNA molecules to be PCR amplified and sequenced in bulk. This technique eliminates re-sampling of multiple clones from the same initial template and in principal the effects of PCR-based recombination (125). However, we found that by producing full length *env* sequences, fragments are generated during RT-PCR, which are able to recombine during further amplification steps (unpublished data, P. Rieder). To avoid Taq polymerase errors and amplification from more than one template, sequences with “double peaks” in the chromatogram has to be excluded (71). This reduces the error rate due to PCR and the effects of PCR-based recombination, but elevates the effort dramatically.

Ultradeep sequencing

The automated Sanger method is considered as a “first generation” technology, and newer methods are referred to as next-generation sequencing (NGS) including ultradeep sequencing. Currently, 5 NGS platforms are commercially available, Roche/454 life science, Illumina/Solexa, Applied Biosystems/SOLiD, Life/APG and Helicos BioSciences. NGS platforms share a common technological feature, massively parallel sequencing of clonally amplified or single DNA molecules that are spatially separated in a flow cell. The major advance offered by NGS is the ability to produce an enormous volume of data cheaply.
The 454 technology (www.454.com) is derived from the technological convergence of pyrosequencing and emulsion PCR (reviewed in (92)). Ultra deep sequencing is a new field which is largely being enabled through 454 Sequencing technology. This method is designed to allow mutations to be detected at extremely low levels.

11. Tests to evaluate HIV-1 coreceptor tropism

Determination of HIV-1 coreceptor usage (figure 7) is becoming a crucial step for optimal therapy with drugs aimed to block the interaction of the virus with CCR5 receptors (155). There are different methodologies to determine the ability of the virus to use one or both coreceptors.

Phenotypic assays to predict HIV-1 coreceptor usage

Original HIV-1 phenotypic classifications used the ability of the virus to cause or not to cause syncytia in cell cultures. MT-2 cells, are used to differentiate between syncytium-inducing (SI) and nonsyncytium-inducing (NSI) HIV-1 variants (65). MT-2 cell lines express CXCR4 but not CCR5 on the cell surface and indicate the presence (SI) or absence (NSI) of X4 viral strains. However, this assay can not determine if dual tropic or R5 variants are present.

More recent methods use HIV-1 isolates to infect reporter cell lines expressing a set of specific receptors. One indicator cell line frequently used is GHOST, which is derived from human osteosarcoma cells and carries an HIV-1 Tat-inducible GFP gene, CD4 receptors and either CCR5 or CXCR4 coreceptors (144).

Both assays use HIV-1 isolates which have a few important caveats. Standard viral isolation procedures require coculturing of the patient's HIV-1 infected PBMCs with PBMCs from HIV-seronegative donors. Extended virus culturing may lead to adaptation of the quasispecies to the new environment (145). Small differences in HIV-1 isolation protocols may alter the level of coreceptor expression on the cell surface (18).

A viable alternative to the use of HIV-1 isolates involves the use of *env*-recombinant viruses and reporter cell lines expressing CD4 and coreceptors. In the commercially available Trofile assay (Monogram Biosciences), the entire *env* is amplified from plasma samples of HIV-infected individuals and cloned into a vector. This vector is contransfected into HEK293 cells with an HIV genomic vector carrying a luciferase reporter gene instead of *env*. These pseudotyped replication-defective viruses are used to infect U87 cells expressing CD4 and either CCR5 or CXCR4. Coreceptor usage is quantified by measuring luciferase activity after a single round of infection (158).

Genotypic assays to predict viral tropism

The first genotypic algorithm designed to predict HIV-1 tropism took into account only the charge of amino acids (AA) at two key residues located within the V3 loop (AA 11 and 25). But most current genotypic bioinformatic tools consider the entire V3 sequence (117). Moreover, positions outside the V3 loop may influence viral tropism as well (104). However, most genotypic predictors have been designed based on HIV-1 clade B sequences (112, 138). Bioinformatic methods that use sequence directly the sequence are cheap and fast, because sequencing has become a routine task in biotechnology.

Today, bioinformatic tools use machine-learning methods to classify the virus with respect to its coreceptor usage, such as support vector machines (SVM) (110, 137) or position-specific scoring matrices (PSSM) (66). The PSSM method assigns a score based on the similarity of sequence to known X4 V3 sequences at each AA site, effectively weighted by the site's ability to discriminate between the two phenotypes. The higher the score, the more likely it is that a V3 sequence comes from an X4 strain. Dual-tropic viruses, thought to be evolutionary intermediates between R5 and X4 virus (23), are also intermediate in PSSM score (67). A support vector machine (SVM) is used for classification and regression analysis. A SVM takes a set of input data and predicts, for each given input, to which of two possible classes the input belongs. SVM uses a set of training examples for classification so that the separate categories are divided by a clear gap that is as wide as possible.

12. Phylogenetic tree

Evolutionary relationships among genes can be illustrated by a phylogenetic tree, comparable to a pedigree showing which genes are most closely related. Terminal nodes are the extant taxa or viral strains, whereas internal nodes are the hypothetical progenitors. A group of taxa that belong to the same branch have a monophyletic origin and is called a cluster.

Reconstructing phylogeny results in an inferred phylogenetic tree, which may or may not differ from the true phylogenetic tree. There are no uniquely correct methods for inferring phylogenies, and many methods are used. The methods can be grouped first according to whether the method uses discrete character states or a distance

matrix of pairwise dissimilarities, and second according to whether the method clusters extant taxa stepwise, resulting in only one best tree, or considers all theoretically possible trees (exhaustive search). The majority of distance-matrix methods use stepwise clustering to compute the best tree, whereas most character-state methods adopt the exhaustive-search approach.

Maximum likelihood (ML) examines every reasonable tree topology and evaluates the support for each by examining every sequence position. In principle, the ML algorithm calculates the probability of expecting each possible nucleotide in the internal nodes and infers the likelihood of the tree structure from these probabilities. The likelihood of all reasonable tree topologies is searched in this way, and the most likely tree is chosen as the best tree. The ML method is complex and computationally demanding (126).

Neighbor-joining (NJ) is based on the minimum-evolution criterion, i.e. the topology that minimizes the length of all internal branches and thus the length of the entire tree. However, neighbor-joining may not find the true tree topology with least total branch length because it is a greedy algorithm that constructs the tree in a step-wise fashion. Even though it is sub-optimal in this sense, it often identifies trees that are quite close to the optimal tree. The main virtue of neighbor-joining relative to these other methods is its computational efficiency (126).

References

1. Abbate, I., C. Vlassi, G. Rozera, A. Bruselles, B. Bartolini, E. Giombini, A. Corpolongo, G. D'Offizi, P. Narciso, A. Desideri, G. Ippolito, and M. R. Capobianchi. 2010. Detection of quasispecies variants predicted to use CXCR4 by ultra-deep pyrosequencing during early HIV infection. AIDS.
2. Abrahams, M. R., J. A. Anderson, E. E. Giorgi, C. Seoighe, K. Mlisana, L. H. Ping, G. S. Athreya, F. K. Treurnicht, B. F. Keele, N. Wood, J. F. Salazar-Gonzalez, T. Bhattacharya, H. Chu, I. Hoffman, S. Galvin, C. Mapanje, P. Kazembe, R. Thebus, S. Fiscus, W. Hide, M. S. Cohen, S. A. Karim, B. F. Haynes, G. M. Shaw, B. H. Hahn, B. T. Korber, R. Swanstrom, and C. Williamson. 2009. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. J Virol 83:3556-67.
3. Aceto, L., U. Karrer, C. Grube, R. Oberholzer, B. Hasse, E. Presterl, J. Boni, H. Kuster, A. Trkola, R. Weber, and H. F. Gunthard. 2005. [Primary HIV-1 infection in Zurich: 2002-2004]. Praxis (Bern 1994) 94:1199-205.
4. Alizon, S., V. von Wyl, T. Stadler, R. D. Kouyos, S. Yerly, B. Hirschel, J. Boni, C. Shah, T. Klimkait, H. Furrer, A. Rauch, P. L. Vernazza, E. Bernasconi, M. Battegay, P. Burgisser, A. Telenti, H. F. Gunthard, and S. Bonhoeffer. 2010. Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. PLoS Pathog 6.
5. Arien, K. K., G. Vanham, and E. J. Arts. 2007. Is HIV-1 evolving to a less virulent form in humans? Nat Rev Microbiol 5:141-51.
6. Auerbach, D. M., W. W. Darrow, H. W. Jaffe, and J. W. Curran. 1984. Cluster of cases of the acquired immune deficiency syndrome. Patients linked by sexual contact. Am J Med 76:487-92.
7. BAG, B. f. G. 2011. HIV-Quartalszahlen per 31. Dezember 2010. Bundesamt für Gesundheit BAG.
8. BAG, B. f. G. 2011, posting date. HIV und AIDS in der Schweiz: Eckdaten per Ende 2010. [Online.]
9. BAG, B. f. G. 2008. HIV/Aids in der Schweiz am 30. September 2008. Bundesamt für Gesundheit BAG.
10. Bar, K. J., H. Li, A. Chamberland, C. Tremblay, J. P. Routy, T. Grayson, C. Sun, S. Wang, G. H. Learn, C. J. Morgan, J. E. Schumacher, B. F. Haynes, B. F. Keele, B. H. Hahn, and G. M. Shaw. 2010. Wide variation in the multiplicity of HIV-1 infection among injection drug users. J Virol 84:6241-7.
11. Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220:868-71.
12. Berger, E. A., P. M. Murphy, and J. M. Farber. 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 17:657-700.
13. Binley, J. M., H. Arshad, T. R. Fouts, and J. P. Moore. 1997. An investigation of the high-avidity antibody response to glycoprotein 120 of human immunodeficiency virus type 1. AIDS Res Hum Retroviruses 13:1007-15.
14. Borrow, P., H. Lewicki, B. H. Hahn, G. M. Shaw, and M. B. Oldstone. 1994. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of

viremia in primary human immunodeficiency virus type 1 infection. J Virol 68:6103-10.
15. Brenner, B. G., M. Roger, J. P. Routy, D. Moisi, M. Ntemgwa, C. Matte, J. G. Baril, R. Thomas, D. Rouleau, J. Bruneau, R. Leblanc, M. Legault, C. Tremblay, H. Charest, and M. A. Wainberg. 2007. High rates of forward transmission events after acute/early HIV-1 infection. J Infect Dis 195:951-9.
16. Brown, A. E., R. J. Gifford, J. P. Clewley, C. Kucherer, B. Masquelier, K. Porter, C. Balotta, N. K. Back, L. B. Jorgensen, C. de Mendoza, K. Bhaskaran, O. N. Gill, A. M. Johnson, and D. Pillay. 2009. Phylogenetic reconstruction of transmission events from individuals with acute HIV infection: toward more-rigorous epidemiological definitions. J Infect Dis 199:427-31.
17. Buonaguro, L., M. L. Tornesello, and F. M. Buonaguro. 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J Virol 81:10209-19.
18. Carroll, R. G., J. L. Riley, B. L. Levine, Y. Feng, S. Kaushal, D. W. Ritchey, W. Bernstein, O. S. Weislow, C. R. Brown, E. A. Berger, C. H. June, and D. C. St Louis. 1997. Differential regulation of HIV-1 fusion cofactor expression by CD28 costimulation of CD4+ T cells. Science 276:273-6.
19. Castonguay, L. A., Y. Weng, W. Adolfsen, J. Di Salvo, R. Kilburn, C. G. Caldwell, B. L. Daugherty, P. E. Finke, J. J. Hale, C. L. Lynch, S. G. Mills, M. MacCoss, M. S. Springer, and J. A. DeMartino. 2003. Binding of 2-aryl-4-(piperidin-1-yl)butanamines and 1,3,4-trisubstituted pyrrolidines to human CCR5: a molecular modeling-guided mutagenesis study of the binding pocket. Biochemistry 42:1544-50.
20. Chohan, B., D. Lang, M. Sagar, B. Korber, L. Lavreys, B. Richardson, and J. Overbaugh. 2005. Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. J Virol 79:6528-31.
21. Christopherson, C., J. Sninsky, and S. Kwok. 1997. The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. Nucleic Acids Res 25:654-8.
22. Condra, J. H., W. A. Schleif, O. M. Blahy, L. J. Gabryelski, D. J. Graham, J. C. Quintero, A. Rhodes, H. L. Robbins, E. Roth, M. Shivaprakash, and et al. 1995. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. Nature 374:569-71.
23. Connor, R. I., K. E. Sheridan, D. Ceradini, S. Choe, and N. R. Landau. 1997. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. J Exp Med 185:621-8.
24. Coombs, R. W., P. S. Reichelderfer, and A. L. Landay. 2003. Recent observations on HIV type-1 infection in the genital tract of men and women. AIDS 17:455-80.
25. Curlin, M. E., R. Zioni, S. E. Hawes, Y. Liu, W. Deng, G. S. Gottlieb, T. Zhu, and J. I. Mullins. 2010. HIV-1 envelope subregion length variation during disease progression. PLoS Pathog 6:e1001228.
26. Damond, F., B. Roquebert, A. Benard, G. Collin, M. Miceli, P. Yeni, F. Brun-Vezinet, and D. Descamps. 2007. Human immunodeficiency virus type 1 (HIV-1) plasma load discrepancies between the Roche COBAS AMPLICOR HIV-1 MONITOR Version 1.5 and the Roche COBAS AmpliPrep/COBAS TaqMan HIV-1 assays. J Clin Microbiol 45:3436-8.

27. Delwart, E. L., H. W. Sheppard, B. D. Walker, J. Goudsmit, and J. I. Mullins. 1994. Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays. J Virol 68:6672-83.
28. Derdeyn, C. A., J. M. Decker, F. Bibollet-Ruche, J. L. Mokili, M. Muldoon, S. A. Denham, M. L. Heil, F. Kasolo, R. Musonda, B. H. Hahn, G. M. Shaw, B. T. Korber, S. Allen, and E. Hunter. 2004. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science 303:2019-22.
29. Egger, M., B. Hirschel, P. Francioli, P. Sudre, M. Wirz, M. Flepp, M. Rickenbach, R. Malinverni, P. Vernazza, and M. Battegay. 1997. Impact of new antiretroviral combination therapies in HIV infected patients in Switzerland: prospective multicentre study. Swiss HIV Cohort Study. Bmj 315:1194-9.
30. Fatkenheuer, G., M. Nelson, A. Lazzarin, I. Konourina, A. I. Hoepelman, H. Lampiris, B. Hirschel, P. Tebas, F. Raffi, B. Trottier, N. Bellos, M. Saag, D. A. Cooper, M. Westby, M. Tawadrous, J. F. Sullivan, C. Ridgway, M. W. Dunne, S. Felstead, H. Mayer, and E. van der Ryst. 2008. Subgroup analyses of maraviroc in previously treated R5 HIV-1 infection. N Engl J Med 359:1442-55.
31. Fiebig, E. W., D. J. Wright, B. D. Rawal, P. E. Garrett, R. T. Schumacher, L. Peddada, C. Heldebrant, R. Smith, A. Conrad, S. H. Kleinman, and M. P. Busch. 2003. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS 17:1871-9.
32. Fisher, M., D. Pao, G. Murphy, G. Dean, D. McElborough, G. Homer, and J. V. Parry. 2007. Serological testing algorithm shows rising HIV incidence in a UK cohort of men who have sex with men: 10 years application. AIDS 21:2309-14.
33. Fleming, D. T., and J. N. Wasserheit. 1999. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex Transm Infect 75:3-17.
34. Fouchier, R. A., M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, and H. Schuitemaker. 1992. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J Virol 66:3183-7.
35. Frankel, A. D., and J. A. Young. 1998. HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 67:1-25.
36. Frost, S. D. W., T. Wrin, D.M. Smith, et al. 2005. Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. Proc Natl Acad Sci U S A 2005; 102: 18514–9.
37. Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397:436-41.
38. Ghys, P. D., K. Fransen, M. O. Diallo, V. Ettiegne-Traore, I. M. Coulibaly, K. M. Yeboue, M. L. Kalish, C. Maurice, J. P. Whitaker, A. E. Greenberg, and M. Laga. 1997. The associations between cervicovaginal HIV shedding, sexually transmitted diseases and immunosuppression in female sex workers in Abidjan, Cote d'Ivoire. AIDS 11:F85-93.
39. Gianella, S., V. von Wyl, M. Fischer, B. Niederoest, M. Battegay, E. Bernasconi, M. Cavassini, A. Rauch, B. Hirschel, P. Vernazza, R. Weber, B. Joos, and H. F. Gunthard. 2011. Effect of early antiretroviral therapy during

primary HIV-1 infection on cell-associated HIV-1 DNA and plasma HIV-1 RNA. Antivir Ther 16:535-545.
40. Gianella, S., V. vonWyl, M. Fischer, B. Niederoest, M. Battegay, E. Bernasconi, M. Cavassini, A. Rauch, B. Hirschel, P. Vernazza, R. Weber, B. Joos, and F. H. Guenthard. 2011. Effect of early antiretroviral therapy during primary HIV-1 infection on cell-associated HIV-1 DNA and plasma HIV-1 RNA. 2011 16:535-545.
41. Gottlieb, G. S., D. C. Nickle, M. A. Jensen, K. G. Wong, J. Grobler, F. Li, S. L. Liu, C. Rademeyer, G. H. Learn, S. S. Karim, C. Williamson, L. Corey, J. B. Margolick, and J. I. Mullins. 2004. Dual HIV-1 infection associated with rapid disease progression. Lancet 363:619-22.
42. Granich, R. M., C. F. Gilks, C. Dye, K. M. De Cock, and B. G. Williams. 2009. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. Lancet 373:48-57.
43. Gray, R. R., J. Parker, P. Lemey, M. Salemi, A. Katzourakis, and O. G. Pybus. 2011. The mode and tempo of hepatitis C virus evolution within and among hosts. BMC Evol Biol 11:131.
44. Gupta, P., K. B. Collins, D. Ratner, S. Watkins, G. J. Naus, D. V. Landers, and B. K. Patterson. 2002. Memory CD4(+) T cells are the earliest detectable human immunodeficiency virus type 1 (HIV-1)-infected cells in the female genital mucosal tissue during HIV-1 transmission in an organ culture system. J Virol 76:9868-76.
45. Haaland, R. E., P. A. Hawkins, J. Salazar-Gonzalez, A. Johnson, A. Tichacek, E. Karita, O. Manigart, J. Mulenga, B. F. Keele, G. M. Shaw, B. H. Hahn, S. A. Allen, C. A. Derdeyn, and E. Hunter. 2009. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PLoS Pathog 5:e1000274.
46. Haase, A. T. 2011. Early events in sexual transmission of HIV and SIV and opportunities for interventions. Annu Rev Med 62:127-39.
47. Hammer, S. M., D. A. Katzenstein, M. D. Hughes, H. Gundacker, R. T. Schooley, R. H. Haubrich, W. K. Henry, M. M. Lederman, J. P. Phair, M. Niu, M. S. Hirsch, and T. C. Merigan. 1996. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. AIDS Clinical Trials Group Study 175 Study Team. N Engl J Med 335:1081-90.
48. Hammer, S. M., K. E. Squires, M. D. Hughes, J. M. Grimes, L. M. Demeter, J. S. Currier, J. J. Eron, Jr., J. E. Feinberg, H. H. Balfour, Jr., L. R. Deyton, J. A. Chodakewitz, and M. A. Fischl. 1997. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. N Engl J Med 337:725-33.
49. Harrington, M., and C. C. Carpenter. 2000. Hit HIV-1 hard, but only when necessary. Lancet 355:2147-52.
50. Hecht, F. M., L. Wang, A. Collier, S. Little, M. Markowitz, J. Margolick, J. M. Kilby, E. Daar, B. Conway, and S. Holte. 2006. A multicenter observational study of the potential benefits of initiating combination antiretroviral therapy during acute HIV infection. J Infect Dis 194:725-33.
51. Hemelaar, J., E. Gouws, P. D. Ghys, and S. Osmanov. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20:W13-23.

52. Herbeck, J. T., D. C. Nickle, G. H. Learn, G. S. Gottlieb, M. E. Curlin, L. Heath, and J. I. Mullins. 2006. Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. J Virol 80:1637-44.
53. Hirsch, M. S., F. Brun-Vezinet, B. Clotet, B. Conway, D. R. Kuritzkes, R. T. D'Aquila, L. M. Demeter, S. M. Hammer, V. A. Johnson, C. Loveday, J. W. Mellors, D. M. Jacobsen, and D. D. Richman. 2003. Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type 1: 2003 recommendations of an International AIDS Society-USA Panel. Clin Infect Dis 37:113-28.
54. Hirsch, M. S., H. F. Gunthard, J. M. Schapiro, F. Brun-Vezinet, B. Clotet, S. M. Hammer, V. A. Johnson, D. R. Kuritzkes, J. W. Mellors, D. Pillay, P. G. Yeni, D. M. Jacobsen, and D. D. Richman. 2008. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. Top HIV Med 16:266-85.
55. Hirsch, V. M., R. A. Olmsted, M. Murphey-Corb, R. H. Purcell, and P. R. Johnson. 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature 339:389-92.
56. Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373:123-6.
57. Hoen, B., D. A. Cooper, F. C. Lampe, L. Perrin, N. Clumeck, A. N. Phillips, L. E. Goh, S. Lindback, D. Sereni, B. Gazzard, J. Montaner, H. J. Stellbrink, A. Lazzarin, D. Ponscarme, S. Staszewski, L. Mathiesen, D. Smith, R. Finlayson, R. Weber, L. Wegmann, G. Janossy, and S. Kinloch-de Loes. 2007. Predictors of virological outcome and safety in primary HIV type 1-infected patients initiating quadruple antiretroviral therapy: QUEST GW PROB3005. Clin Infect Dis 45:381-90.
58. Hoen, B., B. Dumon, M. Harzic, A. Venet, B. Dubeaux, C. Lascoux, Y. Bourezane, J. M. Ragnaud, A. Bicart-See, F. Raffi, L. Beauvais, H. Fleury, and D. Sereni. 1999. Highly active antiretroviral treatment initiated early in the course of symptomatic primary HIV-1 infection: results of the ANRS 053 trial. J Infect Dis 180:1342-6.
59. Hoffmann, Rockstroh, and Kamps. 2007. HIV Medicine.
60. Hollingsworth, T. D., R. M. Anderson, and C. Fraser. 2008. HIV-1 transmission, by stage of infection. J Infect Dis 198:687-93.
61. Huber, M., M. Fischer, B. Misselwitz, A. Manrique, H. Kuster, B. Niederost, R. Weber, V. von Wyl, H. F. Gunthard, and A. Trkola. 2006. Complement lysis activity in autologous plasma is associated with lower viral loads during the acute phase of HIV-1 infection. PLoS Med 3:e441.
62. Huber, M., and A. Trkola. 2007. Humoral immunity to HIV-1: neutralization and beyond. J Intern Med 262:5-25.
63. Hue, S., J. P. Clewley, P. A. Cane, and D. Pillay. 2004. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. AIDS 18:719-28.
64. Janssen, R. S., G. A. Satten, S. L. Stramer, B. D. Rawal, T. R. O'Brien, B. J. Weiblen, F. M. Hecht, N. Jack, F. R. Cleghorn, J. O. Kahn, M. A. Chesney, and M. P. Busch. 1998. New Testing Strategy to Detect Early HIV-1 Infection for Use in Incidence Estimates and for Clinical and Prevention Purposes. JAMA 280:42-48.
65. Japour, A. J., S. A. Fiscus, J. M. Arduino, D. L. Mayers, P. S. Reichelderfer, and D. R. Kuritzkes. 1994. Standardized microtiter assay for determination of

syncytium-inducing phenotypes of clinical human immunodeficiency virus type 1 isolates. J Clin Microbiol 32:2291-4.
66. Jensen, M. A., F. S. Li, A. B. van 't Wout, D. C. Nickle, D. Shriner, H. X. He, S. McLaughlin, R. Shankarappa, J. B. Margolick, and J. I. Mullins. 2003. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol 77:13376-88.
67. Jensen, M. A., and A. B. van 't Wout. 2003. Predicting HIV-1 coreceptor usage with sequence analysis. AIDS Rev 5:104-12.
68. Joos, B., A. Trkola, M. Fischer, H. Kuster, P. Rusert, C. Leemann, J. Boni, A. Oxenius, D. A. Price, R. E. Phillips, J. K. Wong, B. Hirschel, R. Weber, and H. F. Gunthard. 2005. Low human immunodeficiency virus envelope diversity correlates with low in vitro replication capacity and predicts spontaneous control of plasma viremia after treatment interruptions. J Virol 79:9026-37.
69. Karlsson Hedestam, G. B., R. A. Fouchier, S. Phogat, D. R. Burton, J. Sodroski, and R. T. Wyatt. 2008. The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. Nat Rev Microbiol 6:143-55.
70. Kassutto, S., K. Maghsoudi, M. N. Johnston, G. K. Robbins, N. C. Burgett, P. E. Sax, D. Cohen, E. Pae, B. Davis, K. Zachary, N. Basgoz, M. D'Agata E, V. DeGruttola, B. D. Walker, and E. S. Rosenberg. 2006. Longitudinal analysis of clinical markers following antiretroviral therapy initiated during acute or early HIV type 1 infection. Clin Infect Dis 42:1024-31.
71. Keele, B. F., E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L. H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn, and G. M. Shaw. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105:7552-7.
72. Keele, B. F., F. Van Heuverswyn, Y. Li, E. Bailes, J. Takehisa, M. L. Santiago, F. Bibollet-Ruche, Y. Chen, L. V. Wain, F. Liegeois, S. Loul, E. M. Ngole, Y. Bienvenue, E. Delaporte, J. F. Brookfield, P. M. Sharp, G. M. Shaw, M. Peeters, and B. H. Hahn. 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313:523-6.
73. Kempf, D. J., K. C. Marsh, G. Kumar, A. D. Rodrigues, J. F. Denissen, E. McDonald, M. J. Kukulka, A. Hsu, G. R. Granneman, P. A. Baroldi, E. Sun, D. Pizzuti, J. J. Plattner, D. W. Norbeck, and J. M. Leonard. 1997. Pharmacokinetic enhancement of inhibitors of the human immunodeficiency virus protease by coadministration with ritonavir. Antimicrob Agents Chemother 41:654-60.
74. Koup, R. A., J. T. Safrit, Y. Cao, C. A. Andrews, G. McLeod, W. Borkowsky, C. Farthing, and D. D. Ho. 1994. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. J Virol 68:4650-5.
75. Kouyos, R. D., V. von Wyl, S. Yerly, J. Boni, P. Rieder, B. Joos, P. Taffe, C. Shah, P. Burgisser, T. Klimkait, R. Weber, B. Hirschel, M. Cavassini, A. Rauch, M. Battegay, P. L. Vernazza, E. Bernasconi, B. Ledergerber, S. Bonhoeffer, and H. F. Gunthard. 2011. Ambiguous nucleotide calls from

population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection. Clin Infect Dis 52:532-9.
76. Kouyos, R. D., V. von Wyl, S. Yerly, J. Boni, P. Taffe, C. Shah, P. Burgisser, T. Klimkait, R. Weber, B. Hirschel, M. Cavassini, H. Furrer, M. Battegay, P. L. Vernazza, E. Bernasconi, M. Rickenbach, B. Ledergerber, S. Bonhoeffer, and H. F. Gunthard. 2010. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. J Infect Dis 201:1488-97.
77. Kuiken, C., R. Thakallapalli, A. Esklid, and A. de Ronde. 2000. Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus. Am J Epidemiol 152:814-22.
78. Land, A., and I. Braakman. 2001. Folding of the human immunodeficiency virus type 1 envelope glycoprotein in the endoplasmic reticulum. Biochimie 83:783-90.
79. Langford, S. E., J. Ananworanich, and D. A. Cooper. 2007. Predictors of disease progression in HIV infection: a review. AIDS Res Ther 4:11.
80. Larder, B. A., G. Darby, and D. D. Richman. 1989. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. Science 243:1731-4.
81. Lavreys, L., J. M. Baeten, V. Chohan, R. S. McClelland, W. M. Hassan, B. A. Richardson, K. Mandaliya, J. O. Ndinya-Achola, and J. Overbaugh. 2006. Higher set point plasma viral load and more-severe acute HIV type 1 (HIV-1) illness predict mortality among high-risk HIV-1-infected African women. Clin Infect Dis 42:1333-9.
82. Lewis, F., G. J. Hughes, A. Rambaut, A. Pozniak, and A. J. Leigh Brown. 2008. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med 5:e50.
83. Li, H., K. J. Bar, S. Wang, J. M. Decker, Y. Chen, C. Sun, J. F. Salazar-Gonzalez, M. G. Salazar, G. H. Learn, C. J. Morgan, J. E. Schumacher, P. Hraber, E. E. Giorgi, T. Bhattacharya, B. T. Korber, A. S. Perelson, J. J. Eron, M. S. Cohen, C. B. Hicks, B. F. Haynes, M. Markowitz, B. F. Keele, B. H. Hahn, and G. M. Shaw. 2010. High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. PLoS Pathog 6:e1000890.
84. Liu, S. L., A. G. Rodrigo, R. Shankarappa, G. H. Learn, L. Hsu, O. Davidov, L. P. Zhao, and J. I. Mullins. 1996. HIV quasispecies and resampling. Science 273:415-6.
85. Mani, I., P. Gilbert, J. L. Sankale, G. Eisen, S. Mboup, and P. J. Kanki. 2002. Intrapatient diversity and its correlation with viral setpoint in human immunodeficiency virus type 1 CRF02_A/G-IbNG infection. J Virol 76:10745-55.
86. Marcus, U., L. Voss, C. Kollan, and O. Hamouda. 2006. HIV incidence increasing in MSM in Germany: factors influencing infection dynamics. Euro Surveill 11:157-60.
87. Markham, R. B., W. C. Wang, A. E. Weisstein, Z. Wang, A. Munoz, A. Templeton, J. Margolick, D. Vlahov, T. Quinn, H. Farzadegan, and X. F. Yu. 1998. Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. Proc Natl Acad Sci U S A 95:12568-73.
88. Mehandru, S., M. A. Poles, K. Tenner-Racz, A. Horowitz, A. Hurley, C. Hogan, D. Boden, P. Racz, and M. Markowitz. 2004. Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. J Exp Med 200:761-70.
89. Mehandru, S., M. A. Poles, K. Tenner-Racz, P. Jean-Pierre, V. Manuelli, P. Lopez, A. Shet, A. Low, H. Mohri, D. Boden, P. Racz, and M. Markowitz. 2006.

Lack of mucosal immune reconstitution during prolonged treatment of acute and early HIV-1 infection. PLoS Med 3:e484.
90. Mellors, J. W., L. A. Kingsley, C. R. Rinaldo, Jr., J. A. Todd, B. S. Hoo, R. P. Kokka, and P. Gupta. 1995. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. Ann Intern Med 122:573-9.
91. Mellors, J. W., C. R. Rinaldo, Jr., P. Gupta, R. M. White, J. A. Todd, and L. A. Kingsley. 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science 272:1167-70.
92. Metzker, M. L. 2010. Sequencing technologies - the next generation. Nat Rev Genet 11:31-46.
93. Miller, C. J., N. J. Alexander, S. Sutjipto, A. A. Lackner, A. Gettie, A. G. Hendrickx, L. J. Lowenstine, M. Jennings, and P. A. Marx. 1989. Genital mucosal transmission of simian immunodeficiency virus: animal model for heterosexual transmission of human immunodeficiency virus. J Virol 63:4277-84.
94. Miller, C. J., and R. J. Shattock. 2003. Target cells in vaginal HIV transmission. Microbes Infect 5:59-67.
95. Montefiori, D. C., T. S. Hill, H. T. T. Vo, B. D. Walker, E. S. Rosenberg. 2001. Neutralizing antibodies associated with viremia control in a subset of individuals after treatment of acute human immunodeficiency virus type 1 infection. J Virol 2001; 75: 10200–7.
96. Moog, C., H.J. Fleury, I. Pellegrin, A. Kirn, A. M. Aubertin. 1997. Autologous and heterologous neutralizing antibody responses following initial seroconversion in human immunodeficiency virus type 1-infected individuals. J Virol 1997; 71: 3734–41.
97. Nielsen, M. H., F. S. Pedersen, and J. Kjems. 2005. Molecular strategies to inhibit HIV-1 replication. Retrovirology 2:10.
98. Nordt, C., and R. Stohler. 2006. Incidence of heroin use in Zurich, Switzerland: a treatment case register analysis. Lancet 367:1830-4.
99. Oldfield, V., G. M. Keating, and G. Plosker. 2005. Enfuvirtide: a review of its use in the management of HIV infection. Drugs 65:1139-60.
100. Palella, F. J., Jr., K. M. Delaney, A. C. Moorman, M. O. Loveless, J. Fuhrer, G. A. Satten, D. J. Aschman, and S. D. Holmberg. 1998. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. N Engl J Med 338:853-60.
101. Pantazis, N., G. Touloumi, P. Vanhems, J. Gill, H. C. Bucher, and K. Porter. 2008. The effect of antiretroviral treatment of different durations in primary HIV infection. AIDS 22:2441-50.
102. Parren, P.W.J.P. Moore, D.R. Burton, Q. J. Sattentau. 1999. The neutralizing antibody response to HIV-1: viral evasion and escape from humoral immunity. AIDS; 13(Suppl. A): S137–62
103. Parry, J. V., P. P. Mortimer, K. R. Perry, D. Pillay, and M. Zuckerman. 2003. Towards error-free HIV diagnosis: guidelines on laboratory practice. Commun Dis Public Health 6:334-50.
104. Pastore, C., R. Nedellec, A. Ramos, S. Pontow, L. Ratner, and D. E. Mosier. 2006. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. J Virol 80:750-8.
105. Patterson, B. K., A. Landay, J. Andersson, C. Brown, H. Behbahani, D. Jiyamapa, Z. Burki, D. Stanislawski, M. A. Czerniewski, and P. Garcia. 1998. Repertoire of chemokine receptor expression in the female genital tract:

implications for human immunodeficiency virus transmission. Am J Pathol 153:481-90.
106. Perrin, L., L. Kaiser, and S. Yerly. 2003. Travel and the spread of HIV-1 genetic variants. Lancet Infect Dis 3:22-7.
107. Peterlin, B. M., and D. Trono. 2003. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. Nat Rev Immunol 3:97-107.
108. Pilcher, C. D., H. C. Tien, J. J. Eron, Jr., P. L. Vernazza, S. Y. Leu, P. W. Stewart, L. E. Goh, and M. S. Cohen. 2004. Brief but efficient: acute HIV infection and the sexual transmission of HIV. J Infect Dis 189:1785-92.
109. Pilgrim A.K., Pantaleo G., Cohen O.J., et al, 1997. Neutralizing antibody responses to human immunodeficiency virus type 1 in primary infection and long-term-nonprogressive infection. J Infect Dis 1997; 176: 924–32.
110. Pillai, S., B. Good, D. Richman, and J. Corbeil. 2003. A new perspective on V3 phenotype prediction. AIDS Res Hum Retroviruses 19:145-9.
111. Pinkerton, S. D. 2007. How many sexually-acquired HIV infections in the USA are due to acute-phase HIV transmission? AIDS 21:1625-9.
112. Poveda, E., V. Briz, M. Quinones-Mateu, and V. Soriano. 2006. HIV tropism: diagnostic tools and implications for disease progression and treatment with entry inhibitors. AIDS 20:1359-67.
113. Preston, B. D., and J. P. Dougherty. 1996. Mechanisms of retroviral mutation. Trends Microbiol 4:16-21.
114. Preston, B. D., B. J. Poiesz, and L. A. Loeb. 1988. Fidelity of HIV-1 reverse transcriptase. Science 242:1168-71.
115. Quinn, T. C., M. J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M. O. Meehan, T. Lutalo, and R. H. Gray. 2000. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med 342:921-9.
116. Raymond, S., P. Delobel, M. Mavigner, M. Cazabat, S. Encinas, C. Souyris, P. Bruel, K. Sandres-Saune, B. Marchou, P. Massip, and J. Izopet. 2010. CXCR4-using viruses in plasma and peripheral blood mononuclear cells during primary HIV-1 infection and impact on disease progression. AIDS 24:2305-12.
117. Resch, W., N. Hoffman, and R. Swanstrom. 2001. Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. Virology 288:51-62.
118. Ritola, K., C. D. Pilcher, S. A. Fiscus, N. G. Hoffman, J. A. Nelson, K. M. Kitrinos, C. B. Hicks, J. J. Eron, Jr., and R. Swanstrom. 2004. Multiple V1/V2 env variants are frequently present during primary infection with human immunodeficiency virus type 1. J Virol 78:11208-18.
119. Ross, H. A., and A. G. Rodrigo. 2002. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. J Virol 76:11715-20.
120. Rusert, P., A. Krarup, C. Magnus, O. F. Brandenburg, J. Weber, A. Ehlert, R. R. Regoes, H. Günthard, and A. Trkola. 2011. Interaction of the V1V2 loop with a neighboring gp120 unit shields the HIV envelope trimer against cross-neutralizing antibodies. J Exp Med in press.
121. Rusert, P., H. Kuster, B. Joos, B. Misselwitz, C. Gujer, C. Leemann, M. Fischer, G. Stiegler, H. Katinger, W. C. Olson, R. Weber, L. Aceto, H. F. Gunthard, and A. Trkola. 2005. Virus Isolates during Acute and Chronic Human Immunodeficiency Virus Type 1 Infection Show Distinct Patterns of Sensitivity to Entry Inhibitors. J. Virol. 79:8454-8469.

122. Sagar, M., O. Laeyendecker, S. Lee, J. Gamiel, M. J. Wawer, R. H. Gray, D. Serwadda, N. K. Sewankambo, J. C. Shepherd, J. Toma, W. Huang, and T. C. Quinn. 2009. Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. J Infect Dis 199:580-9.
123. Sagar, M., L. Lavreys, J. M. Baeten, B. A. Richardson, K. Mandaliya, B. H. Chohan, J. K. Kreiss, and J. Overbaugh. 2003. Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. J Virol 77:12921-6.
124. Sagar, M., L. Lavreys, J. M. Baeten, B. A. Richardson, K. Mandaliya, J. O. Ndinya-Achola, J. K. Kreiss, and J. Overbaugh. 2004. Identification of modifiable factors that affect the genetic diversity of the transmitted HIV-1 population. Aids 18:615-9.
125. Salazar-Gonzalez, J. F., E. Bailes, K. T. Pham, M. G. Salazar, M. B. Guffey, B. F. Keele, C. A. Derdeyn, P. Farmer, E. Hunter, S. Allen, O. Manigart, J. Mulenga, J. A. Anderson, R. Swanstrom, B. F. Haynes, G. S. Athreya, B. T. Korber, P. M. Sharp, G. M. Shaw, and B. H. Hahn. 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol 82:3952-70.
126. Salemi, M., and A. Vandamme. 2006. the phylogenetic handbook. Cambridge University press.
127. Schacker, T., A. C. Collier, J. Hughes, T. Shea, and L. Corey. 1996. Clinical and epidemiologic features of primary HIV infection. Ann Intern Med 125:257-64.
128. Schacker, T., S. Little, E. Connick, K. Gebhard-Mitchell, Z. Q. Zhang, J. Krieger, J. Pryor, D. Havlir, J. K. Wong, D. Richman, L. Corey, and A. T. Haase. 2000. Rapid accumulation of human immunodeficiency virus (HIV) in lymphatic tissue reservoirs during acute and early HIV infection: implications for timing of antiretroviral therapy. J Infect Dis 181:354-7.
129. Schmid, A., S. Gianella, V. von Wyl, K. J. Metzner, A. U. Scherrer, B. Niederost, C. F. Althaus, P. Rieder, C. Grube, B. Joos, R. Weber, M. Fischer, and H. F. Gunthard. 2010. Profound depletion of HIV-1 transcription in patients initiating antiretroviral therapy during acute infection. PLoS ONE 5:e13310.
130. Schmitz, J. E., M. J. Kuroda, S. Santra, M. A. Simon, M. A. Lifton, W. Lin, R. Khunkhun, M. Piatak, J. D. Lifson, G. Grosschupff, R. S. Gelman, P. Racz, K. Tenner-Racz, K. A. Mansfield, N. L. Letvin, D. C. Montefiori, and K. A. Reimann. 2003. Effect of humoral immune responses on controlling viremia during primary infection of rhesus monkeys with simian immunodeficiency virus. J Virol 77:2165-73.
131. Schoeni-Affolter, F., B. Ledergerber, M. Rickenbach, C. Rudin, H. F. Gunthard, A. Telenti, H. Furrer, S. Yerly, and P. Francioli. 2010. Cohort profile: the Swiss HIV Cohort study. Int J Epidemiol 39:1179-89.
132. Shankarappa, R., J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, C. R. Rinaldo, G. H. Learn, X. He, X. L. Huang, and J. I. Mullins. 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol 73:10489-502.
133. Sharp, P. M., and B. H. Hahn. 2008. AIDS: prehistory of HIV-1. Nature 455:605-6.
134. Shattock, R. J., and J. P. Moore. 2003. Inhibiting sexual transmission of HIV-1 infection. Nat Rev Micro 1:25-34.

135. SHCS May 2010 2010, posting date. Swiss HIV Cohort Study: State of the Cohort. [Online.]
136. Sheth, P. M., C. Kovacs, K. S. Kemal, R. B. Jones, J. M. Raboud, R. Pilon, C. la Porte, M. Ostrowski, M. Loutfy, H. Burger, B. Weiser, and R. Kaul. 2009. Persistent HIV RNA shedding in semen despite effective antiretroviral therapy. AIDS 23:2050-4.
137. Sing, T., A. J. Low, N. Beerenwinkel, O. Sander, P. K. Cheung, F. S. Domingues, J. Buch, M. Daumer, R. Kaiser, T. Lengauer, and P. R. Harrigan. 2007. Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. Antivir Ther 12:1097-106.
138. Skrabal, K., A. J. Low, W. Dong, T. Sing, P. K. Cheung, F. Mammano, and P. R. Harrigan. 2007. Determining human immunodeficiency virus coreceptor use in a clinical setting: degree of correlation between two phenotypic assays and a bioinformatic model. J Clin Microbiol 45:279-84.
139. Smith, D. E., B. D. Walker, D. A. Cooper, E. S. Rosenberg, and J. M. Kaldor. 2004. Is antiretroviral treatment of primary HIV infection clinically justified on the basis of current evidence? AIDS 18:709-18.
140. Speck, C. E., R. W. Coombs, L. A. Koutsky, J. Zeh, S. O. Ross, T. M. Hooton, A. C. Collier, L. Corey, A. Cent, J. Dragavon, W. Lee, E. J. Johnson, R. R. Sampoleo, and J. N. Krieger. 1999. Risk factors for HIV-1 shedding in semen. Am J Epidemiol 150:622-31.
141. Starcich, B. R., B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, R. C. Gallo, and et al. 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. Cell 45:637-48.
142. Sullivan, P. S., O. Hamouda, V. Delpech, J. E. Geduld, J. Prejean, C. Semaille, J. Kaldor, C. Folch, E. Op de Coul, U. Marcus, G. Hughes, C. P. Archibald, F. Cazein, A. McDonald, J. Casabona, A. van Sighem, and K. A. Fenton. 2009. Reemergence of the HIV epidemic among men who have sex with men in North America, Western Europe, and Australia, 1996-2005. Ann Epidemiol 19:423-31.
143. Thompson, M. A., J. A. Aberg, P. Cahn, J. S. Montaner, G. Rizzardini, A. Telenti, J. M. Gatell, H. F. Gunthard, S. M. Hammer, M. S. Hirsch, D. M. Jacobsen, P. Reiss, D. D. Richman, P. A. Volberding, P. Yeni, and R. T. Schooley. 2010. Antiretroviral treatment of adult HIV infection: 2010 recommendations of the International AIDS Society-USA panel. JAMA 304:321-33.
144. Trkola, A., T. Ketas, V. N. Kewalramani, F. Endorf, J. M. Binley, H. Katinger, J. Robinson, D. R. Littman, and J. P. Moore. 1998. Neutralization sensitivity of human immunodeficiency virus type 1 primary isolates to antibodies and CD4-based reagents is independent of coreceptor usage. J Virol 72:1876-85.
145. Trouplin, V., F. Salvatori, F. Cappello, V. Obry, A. Brelot, N. Heveker, M. Alizon, G. Scarlatti, F. Clavel, and F. Mammano. 2001. Determination of coreceptor usage of human immunodeficiency virus type 1 from patient plasma samples by using a recombinant phenotypic assay. J Virol 75:251-9.
146. UNAIDS 2011, posting date. Report on the global AIDS epidemic 2010. [Online.]
147. Van Heuverswyn, F., Y. Li, C. Neel, E. Bailes, B. F. Keele, W. Liu, S. Loul, C. Butel, F. Liegeois, Y. Bienvenue, E. M. Ngolle, P. M. Sharp, G. M. Shaw, E. Delaporte, B. H. Hahn, and M. Peeters. 2006. Human immunodeficiency viruses: SIV infection in wild gorillas. Nature 444:164.

148. Veazey, R. S., M. DeMaria, L. V. Chalifoux, D. E. Shvetz, D. R. Pauley, H. L. Knight, M. Rosenzweig, R. P. Johnson, R. C. Desrosiers, and A. A. Lackner. 1998. Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. Science 280:427-31.
149. Verhoeven, D., S. Sankaran, M. Silvey, and S. Dandekar. 2008. Antiviral therapy during primary simian immunodeficiency virus infection fails to prevent acute loss of CD4+ T cells in gut mucosa but enhances their rapid restoration through central memory T cells. J Virol 82:4016-27.
150. Vernazza P, Hirschel B, Bernasconi E, and F. M. 2008. Les personnes seropositives ne souff rant d'aucune autre MST et suivant un traitment antiretroviral effi cie ne transmettent pas le VIH par voie sexuelle. *Bull Med Suisses* 89:165-169.
151. Von Wyl, V., R. D. Kouyos, S. Yerly, J. Böni, C. Shah, P. Bürgisser, T. Klimkait, R. Weber, B. Hirschel, M. Cavassini, C. Staehelin, M. Battegay, P. Vernazza, E. Bernasconi, B. Ledergerber, S. Bonhoeffer, and H. F. Günthard. 2011. The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. JID in press.
152. von Wyl, V., S. Yerly, J. Boni, P. Burgisser, T. Klimkait, M. Battegay, H. Furrer, A. Telenti, B. Hirschel, P. L. Vernazza, E. Bernasconi, M. Rickenbach, L. Perrin, B. Ledergerber, and H. F. Gunthard. 2007. Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. Arch Intern Med 167:1782-90.
153. Wang, S., F. Xu, and U. Demirci. 2010. Advances in developing HIV-1 viral load assays for resource-limited settings. Biotechnol Adv 28:770-81.
154. Wawer, M. J., R. H. Gray, N. K. Sewankambo, D. Serwadda, X. Li, O. Laeyendecker, N. Kiwanuka, G. Kigozi, M. Kiddugavu, T. Lutalo, F. Nalugoda, F. Wabwire-Mangen, M. P. Meehan, and T. C. Quinn. 2005. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J Infect Dis 191:1403-9.
155. Weber, J., H. Piontkivska, and M. E. Quinones-Mateu. 2006. HIV type 1 tropism and inhibitors of viral entry: clinical implications. AIDS Rev 8:60-77.
156. Wei, X., J. M. Decker, S. Wang, H. Hui, J. C. Kappes, X. Wu, J. F. Salazar-Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. L. Komarova, M. A. Nowak, B. H. Hahn, P. D. Kwong, and G. M. Shaw. 2003. Antibody neutralization and escape by HIV-1. Nature 422:307-12.
157. Wei, X., S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, and et al. 1995. Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373:117-22.
158. Whitcomb, J. M., W. Huang, S. Fransen, K. Limoli, J. Toma, T. Wrin, C. Chappey, L. D. Kiss, E. E. Paxinos, and C. J. Petropoulos. 2007. Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. Antimicrob Agents Chemother 51:566-75.
159. Wolfs, T. F., G. Zwart, M. Bakker, and J. Goudsmit. 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology 189:103-10.
160. Wolinsky, S. M., B. T. Korber, A. U. Neumann, M. Daniels, K. J. Kunstman, A. J. Whetsell, M. R. Furtado, Y. Cao, D. D. Ho, and J. T. Safrit. 1996. Adaptive

evolution of human immunodeficiency virus-type 1 during the natural course of infection. Science 272:537-42.
161. Wolinsky, S. M., C. M. Wike, B. T. Korber, C. Hutto, W. P. Parks, L. L. Rosenblum, K. J. Kunstman, M. R. Furtado, and J. L. Munoz. 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science 255:1134-7.
162. Wong, J. K., M. C. Strain, R. Porrata, E. Reay, S. Sankaran-Walters, C. C. Ignacio, T. Russell, S. K. Pillai, D. J. Looney, and S. Dandekar. 2010. In vivo CD8+ T-cell suppression of siv viremia is not mediated by CTL clearance of productively infected cells. PLoS Pathog 6:e1000748.
163. Worobey, M., M. Gemmel, D. E. Teuwen, T. Haselkorn, K. Kunstman, M. Bunce, J. J. Muyembe, J. M. Kabongo, R. M. Kalengayi, E. Van Marck, M. T. Gilbert, and S. M. Wolinsky. 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455:661-4.
164. Wyatt, R., and J. Sodroski. 1998. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. Science 280:1884-8.
165. Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. M. Sharp, and D. D. Ho. 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 391:594-7.
166. Zhu, T., H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup, and D. D. Ho. 1993. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. Science 261:1179-81.

CHAPTER 1

Characterization of HIV-1 Diversity and Tropism in 145 Patients with Primary HIV-1 Infection.

Published as Rieder *et al*., CID 2011, in revision

Own contribution

I participated in data acquisition, more precisely in *env* C2-V3-C3 clonal sequencing and full length *env* single genome sequencing, and analysis and interpretation of the sequence traces by bioinformatic tools. I did also the phylogenetic analysis using different bioinformatic tools. Further, I participated in acquisition, interpretation and analysis of clinical data. I also did the statistic analysis and interpretation of the data. I prepared the manuscript and was involved in review and revision of the manuscript and approval prior to submission.

Characterization of HIV-1 Diversity and Tropism in 145 Patients with Primary HIV-1 Infection.

Philip Rieder[1], Beda Joos[1], Alexandra U. Scherrer[1], Herbert Kuster[1], Dominique Braun[1], Christina Grube[1], Barbara Niederöst[1], Christine Leemann[1], Sara Gianella[1], Karin J. Metzner[1], Jürg Böni[2], Rainer Weber[1], Huldrych F. Günthard[1]

[1] Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich, Switzerland;

[2] Swiss National Center for Retroviruses, Institute of Medical Virology, University of Zurich, Switzerland,

Abstract

Background. In the context of sexual transmission of human immunodeficiency virus type 1 (HIV-1) current findings suggest that the mucosal barrier is the major site of viral selection transforming the complex inoculum to a small, homogeneous founder virus population. We analyzed HIV-1 transmission in relation to viral and host characteristics within the Zurich primary HIV-1 infection (ZPHI)–study.

Methods. Clonal HIV-1 envelope sequences (on average 16 clones/patient) were isolated from the first available plasma samples during the early phase of infection from 145 patients with primary HIV-1 infection (PHI). Phylogenetic and tropism analyses were performed. Differences of viral diversities were investigated in association with several parameters potentially influencing HIV-1 transmission, e.g., concomitant sexually transmitted infections (STI) and mode of transmission.

Results. Median viral diversity within *env* C2-V3-C3 region was 0.39% (range 0.04-3.23%). Viral diversity did not correlate with viral load, but it was slightly correlated with the duration of infection. Neither transmission mode, gender, nor STI predicted transmission of more heterogeneous founder virus populations that were found in 16/145 patients (11%; diversity >1%). Only two patients (1.4%) was assuredly infected with CXCR4-tropic HIV-1 within a R5/X4-tropic mixed population as revealed and confirmed using several genotypic prediction algorithms and phenotypic assays.

Conclusions: Our findings suggest that transmission of multiple HIV-1 variants might be a complex process not dependent on mucosal factors alone. CXCR4-tropic viruses can be sexually transmitted in rare instances, but their clinical relevance remains to be determined.

INTRODUCTION

The extensive genetic diversity of HIV-1 is a tremendous challenge regarding the development of broadly effective vaccines and also for antiretroviral treatment. The genetic bottleneck during HIV-1 transmission may be an Achilles heel of HIV-1 [1]. The mucosal barrier could be a key factor in driving the genetically complex viral inoculum to a homogeneous founder population. The impact of the mucosal barrier may depend on the anatomy, physiology, concurrent sexual transmitted infections (STI), and sexual practice [2-4]. However, a genetic bottleneck has also been observed in intravenous drug users (IVDU) indicating mechanisms independent of the mucosa [3, 5].

During transmission, there is also strong selection for viral variants using CCR5 as a co-receptor [6, 7]. In depth characterization of HIV-1 tropism in larger groups of patients during primary HIV-1 infection (PHI) is needed to revisit recent work reporting varying frequencies of CXCR4-tropic strains during PHI [8-10]. Such data is of particular importance, because CCR5 antagonists are approved for clinical use mainly in salvage treatment [11, 12].

Here, we analyzed the complexity of virus populations within the C2-V3-C3 region of the HIV-1 envelope in 145 acutely and recently infected ZPHI-study patients [13-17] belonging to different transmission groups and infected with various viral subtypes. We investigated viral and host characteristics associated with elevated viral diversity and co-receptor tropism.

MATERIALS AND METHODS

Patient characteristics

Patients were enrolled in the Zurich Primary HIV-1 Infection study (ZPHI, http://clinicaltrials.gov, ID=NCT00537966) [13-17]. Acute/recent PHI was confirmed in all patients according to previously published definitions [15]. We estimated a date of infection for each patient by integrating all available clinical and laboratory data [15, 18]. During the first visit, each patient was checked for symptoms and physical signs of concurrent STIs and serology for syphilis. When dysuria, genital ulcers, groin, and/or rectal pain was present, urine and/or rectal swabs were subjected to PCR analysis for Neisseria gonorrhoeae, Chlamydia trachomatis (when positive, specific PCR for lymphogranuloma venereum associated L-serovars was performed), and Herpes simplex.

Sequencing

RNA extraction, amplification, cloning, and sequencing of HIV-1 *env* C2–V3–C3 fragments were performed according to [15] by modification of previously described methods [19, 20].

Genotypic prediction of HIV-1 co-receptor usage

V3-loop sequences were interpreted by three different genotypic prediction tools: 1) Web-PSSM (http://indra.mullins.microbiol.washington.edu/webpssm/), 2) Wetcat **(**http://genomiac2.ucsd.edu:8080/wetcat/v3.html), and 3) geno2pheno[coreceptor] [7] (http://coreceptor.bioinf.mpi-inf.mpg.de/index.php**)**. We used geno2pheno[coreceptor] with a false-positive rate of 5.75% [21]. Wetcat analysis is based on SVM classifier.

PSSM predictions were performed by HIV-1 subtype B and C sinsi matrices for all non-C subtypes and clade C, respectively.

Phenotyping assays

The cell lines MT-2 (D. Richman [22]), GHOST Parental Cell Line, and GHOST Cell Transfectants GHOST CXCR4 and Hi-5 (VN. KewalRamani and DR. Littman [23]) were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH. Syncytium-inducing (SI) phenotypes were determined in MT-2 cell culture assays [24]. If available, primary isolates derived from both, plasma and PBMCs, were tested. In addition, GHOST cell lines expressing either CCR5 or CXCR4 and CD4 or CD4 alone were used to determine co-receptor usage. Cells were infected with primary virus isolates of each patient at MOI=0.06 (multiplicity of infection). Expression of GFP under control of the HIV-2 long terminal repeat promoter was detected by fluorescence microscopy [25].

Phylogenetic analyses

Sequences were edited with SeqMan-5.08 software (DNASTAR Inc., Madison, WI), aligned with MAFFT-6.240 [26, 27], manually corrected, and tested for hypermutations by Hypermut 2.0 (www.hiv.lanl.gov). Molecular evolutionary analyses were conducted using MEGA-4 [28]. Neighbour joining phylogenetic trees were constructed by MEGA-4 as well as seqboot, dnadist, and 'neighbour and consense' (PHYLIP-3.68, distributed by J. Felsenstein). The reference strain HIV-1_{HXB2} (GenBank accession no. K03455) and other B and non-B strains were used as outgroup references and bootstrapping (1,000 or 100 replications with MEGA-4 and PHYLIP-3.68, respectively). *Pol* sequences were obtained from the Swiss HIV Cohort study (SHCS) drug resistance database [29]. Pairwise distances were computed by

using MEGA-4 [28]. Nucleotide diversities were obtained using the Tamura-Nei model. All reported sequences have been deposited in GenBank under accession numbers GU471280 to GU471390, GU471407 to GU471578 and JF958169 to JF960135.

Statistical Analyses

Statistical analyses were performed using GraphPad Prism version 5 (GraphPad Software, San Diego, CA) and STATA 11 SE (StataCorp, College Station, TX). Nonparametric tests were used for group comparison (Mann-Whitney test). Multiple linear regression was used to analyze the association of viral diversity with viral and host baseline characteristics. The following variables were considered in the model: Age, sex, transmission category, STI, antiretroviral syndrome (ARS), viral load, CD4-cell count, viral subtype, and estimated date of infection (EDI).

RESULTS

Patient characteristics

We analyzed 145 patients comprising 131 males and 14 females who were enrolled in the ongoing ZPHI-study (table 1). The modes of transmission according to patients' statements included homosexual (73%), heterosexual (22%), bisexual (2%), IVDU (1.4%), and others (1.4%). Concomitant STIs have been diagnosed in 20 patients (14%). HIV-1 subtype B was most prevalent (80%) followed by CRF01_AE (9%), C (3%), and A (3%). The remaining 5% of patients were infected with HIV-1 subtypes F1, G, CRF02_AG, and CFR12_BF.

125 patients were diagnosed during documented acute HIV-1 infection. Of those, 120 patients (96%) had an acute retroviral syndrome (ARS) [13] and 87 (70%) had a negative or indeterminate Western blot (WB). Recent infection was diagnosed in 20 patients. Baseline blood samples were available within a median of 6 weeks (range 2-18) after estimated date of infection (EDI) in acute and within 12 (range 7-24) weeks in recently infected patients. Fiebig staging [30] was possible for 117 patients: 2 patients were assigned to stage II, 14 patients to stage III, 52 patients to stage IV, 20 patients to stage V, 19 patients to stage V/VI, and 10 patients to stage VI.

Table 1. Baseline characteristics of 145 patients with primary HIV-1 infection

	total patients		female		acute infection		recent
	n	%	n	%	n	%	n
number of patients	145	100	14	100	125	100	20
male	131	90			112	90	19
female	14	10			13	10	1
HIV-1 subtype B [a]	116	80	5	36	99	79	17
transmission mode							
homosexual	106	73			91	73	15
heterosexual	32	22	14	100	28	22	4
bisexual	3	2			2	2	1
IVDU	2	1.4			2	2	
others [b]	2	1.4			2	2	
sexually transmitted infections (STIs) [c]	20	14	1	7	14	11	6
HIV-1 drug resistance transmitted [d]	6	4	1	7	6	5	
acute retroviral syndrom (ARS)	136	94	13	93	120	96	16
negative or indeterminate Western Blot	87	60	6	43	87	70	
	median	(min-max)	median	(min-max)	median	(min-max)	median
estimated duration of infection (weeks)	6	(2 - 24)	6	(2 - 14)	6	(2 - 18)	12
viral load (log_{10} copies of HIV-1 RNA / ml plasma)	5.3	(2.4 - 7.6)	4.8	(4.2 - 6.7)	5.5	(2.4 - 7.6)	5.0
CD4+ T-cells (cells / µl blood)	414	(87 - 1295)	454	(87 - 1295)	374	(127 - 1295)	503.5
age (years)	36	(19 - 70)	32	(19 - 55)	37	(19 - 70)	34

[a] other subtypes: CRF01_AE, C, A, F1, G, CRF02_AE, CRF12_BF

[b] one case needle stick; one case either IVDU or heterosexual

[c] concomitant sexually transmitted infections: syphilis and/or chlamydia and/or gonorrhea

[d] IAS-USA mutation

Phylogenetic reconstruction

Neighbour joining phylogenetic (NJ) trees containing clonal C2-V3-C3 sequences from 113 (78%) patients formed individual clusters with bootstrap-values of 100%. Clusters with bootstrap-values <100% were found in 32 patients, most of whom harboured also closely related *pol* sequences (median genetic distance 0.0062; range 0–0.0124) except for 4 patients forming 2 clusters. Possible contamination was ruled out by analyzing an independent sample from each of these 4 patients. Sequences from all patients showed no clustering with reference clones used in our laboratory.

Validity of viral diversity at baseline

In total, 2268 clones (median 16 clones/patient; range 10-16) spanning C2-V3-C3 region of the *env* gene were derived from plasma HIV-1 RNA of 145 PHI-patients early after transmission (median 6 weeks, range 2-24). The baseline viral diversities ranged from 0.04% to 3.23% (median 0.39%). To obtain representative samples of quasispecies, RT-PCR was performed in duplicate using HIV-1 RNA extracts from 1ml plasma. The viral load of the plasma samples ranged from 2.4-7.6 (median 5.3) $\log_{10}$ copies/ml. Despite this broad distribution, viral loads and nucleotide diversities were not correlated (figure 1). The viral loads of ten patients were <5'000 copies/ml, however, viral diversities (median 0.44%; range 0.11-1.25%) were similar to patients with high viral loads (p=0.662). Thus, the sequenced clones were representative of the actual plasma virus populations *in vivo* as shown before in chronically infected patients [20]. Additionally, single genome amplification (SGA, median 12 sequences/sample) was performed in a subset of 25 patients and the resulting viral diversities correlated well with clonal sequencing (slope=1.05, intercept=-0.28%, r^2=0.681; data not shown). The viral diversities increased slightly with time since EDI

(EDI was described in previous published methods [15, 18]) (figure 2) supporting the reliability of the used EDI calculation. As also shown in figure 2, negative WBs have been detected up to 10 weeks and positive WBs between 4 and 24 weeks after EDI demonstrating the need to include clinical data into the diagnostic staging process.

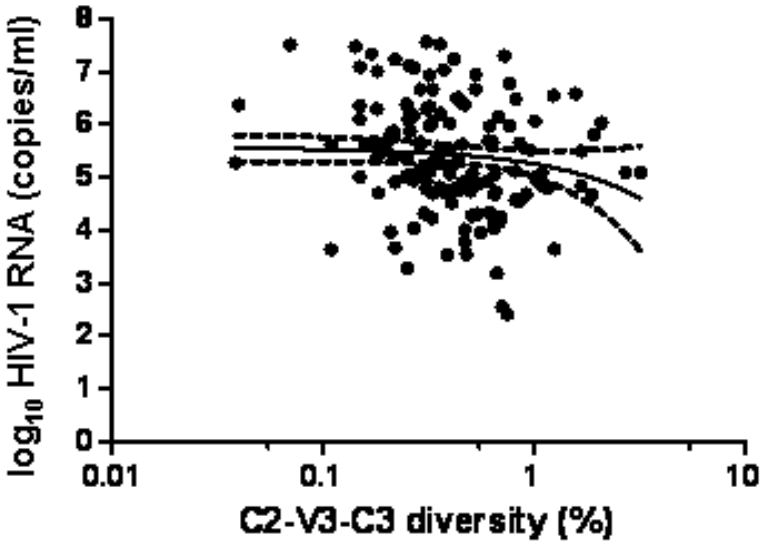

Figure 1. Relationship between plasma viral load and viral diversity during primary HIV-1 infection. Plasma viral load of the first available sample and the corresponding viral nucleotide sequence diversity observed in the C2-V3-C3 region of the envelope gene in 10 to 16 clones per patient is shown. 145 patients have been analyzed. Each dot represents one patient. The relationship between viral diversity and viral load is shown by linear regression ($p=0.109$; slope=-0.299 ± 0.185; $r^2=0.018$).

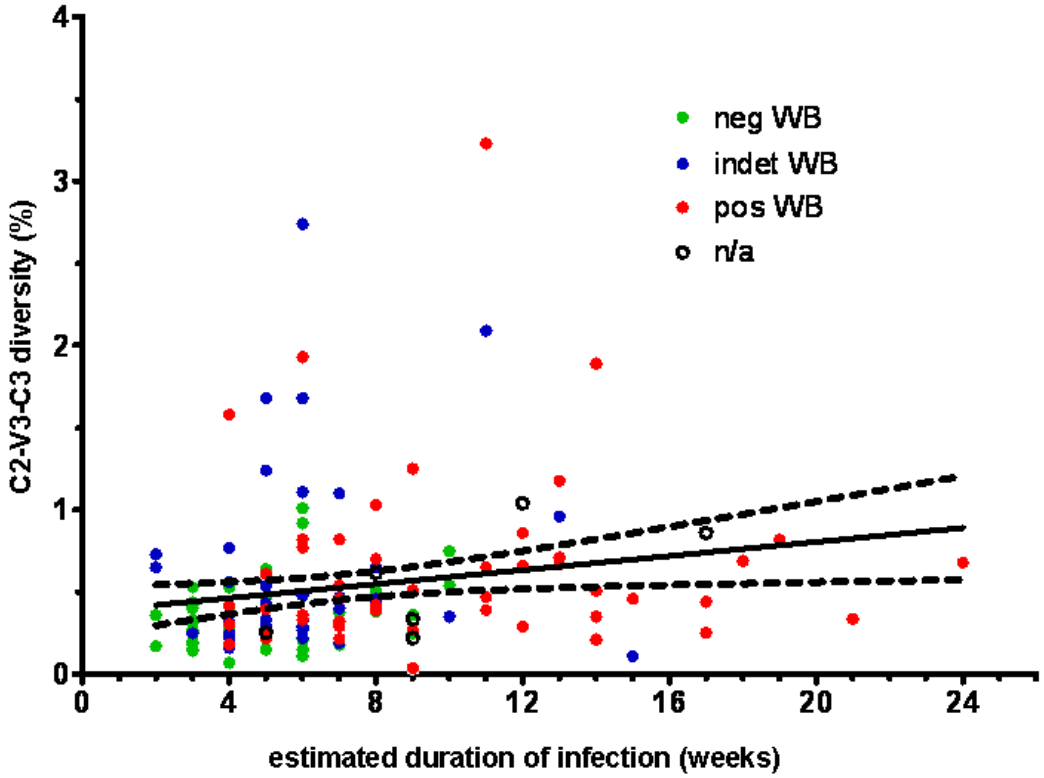

Figure 2. Relationship between viral diversity during primary HIV-1 infection and estimated duration of infection. For each of the 145 patients, the duration between estimated date of infection and first available sample and the corresponding viral diversity observed in the C2-V3-C3 region of the envelope gene in 10 to 16 clones is shown. The duration of infection has been estimated integrating all available information including patient history relating to known risk situations, occurrence of first symptoms, previous negative test results, avidity assays, and Western blot. The median viral diversities increased slightly with longer time intervals between infection and sample collection (slope=0.024±0.009; p=0.015; r^2=0.041). Negative Western blots (WBs) were detected up to 10 weeks and positive WBs between 4 and 24 weeks of EDI, respectively. Neg WB, negative Western blot; indet WB, indeterminate Western blot; pos WB, positive Western blot; n/a, not available.

Viral diversity in relation to host and viral characteristics

In our study population the median viral diversity shortly after transmission did not deviate between different sexual transmission categories (figure 3A). Twenty patients, including 1 female, 5 MSF (men having sex with females), and 14 MSM (men having sex with men), were diagnosed with a concomitant STI, but their median viral diversity did not deviate from those without concomitant STI (p=0.703, Fig. 3B). Next, we analysed in more detail the relation between diversity and transmission mode among MSM depending on the encountered mucosa. Analyzed sexual practises were insertive anal intercourse (iai), receptive anal intercourse (rai), and unsafe oral intercourse (oral). Patients practising various sexual activities or with unknown transmission route were excluded from this analysis. No significant differences between distinct transmission modes were observed (figure 3C).

We also derived the viral subtypes from individual *pol* sequences to examine their possible effect on diversity (figure 4). Four patients have been infected with HIV-1 subtype C. They showed a median viral diversity of 1.07% (range 0.54-1.89%) which significantly deviated from those observed in HIV-1 subtype B (p=0.011). In addition, no association between viral diversity and viral tropism was found (for tropism see below). Moreover, multiple linear regression testing the association of viral diversity with EDI, age, CD4-cell count, viral load, viral subtype, ARS, STI, and transmission mode showed no relationship except for EDI (slope=0.031 [95%-CI=0.001-0.061], p=0.046, r^2=0.151).

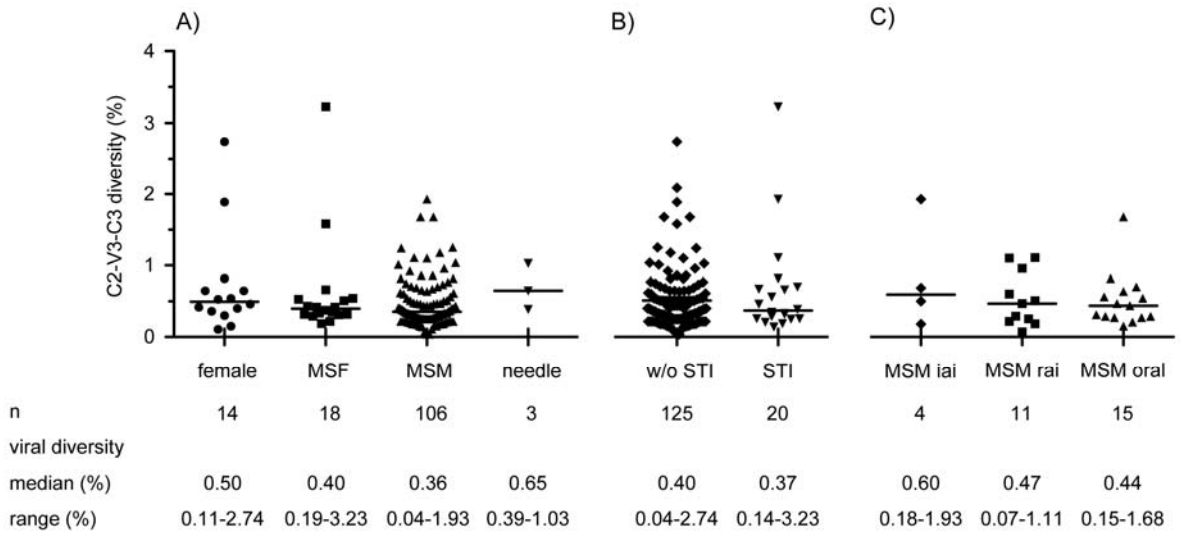

Figure 3. Influence of transmission characteristics on viral diversity during primary HIV-1 infection. Viral diversities observed in the C2-V3-C3 region of the envelope gene in 144 primary HIV-1 infected patients are shown grouped according to gender and transmission category in A) (one patient infected by either sexual intercourse or needle sharing is not shown), according to the presence or absence of a concomitant sexual transmitted infection in B), and within the transmission category "men having sex with men (MSM) according to the sexual practice during HIV-1 acquisition in C). In the latter analysis, MSM patients with unknown transmission mode or practicing several sexual behaviours were excluded from this analysis. Bars indicate medians; MSF, men having sex with females; MSM, men having sex with men; w/o STI, without concomitant sexually transmitted infections; STI, with sexually transmitted infections; iai, insertive anal intercourse; rai, receptive anal intercourse, oral, unsafe oral intercourse.

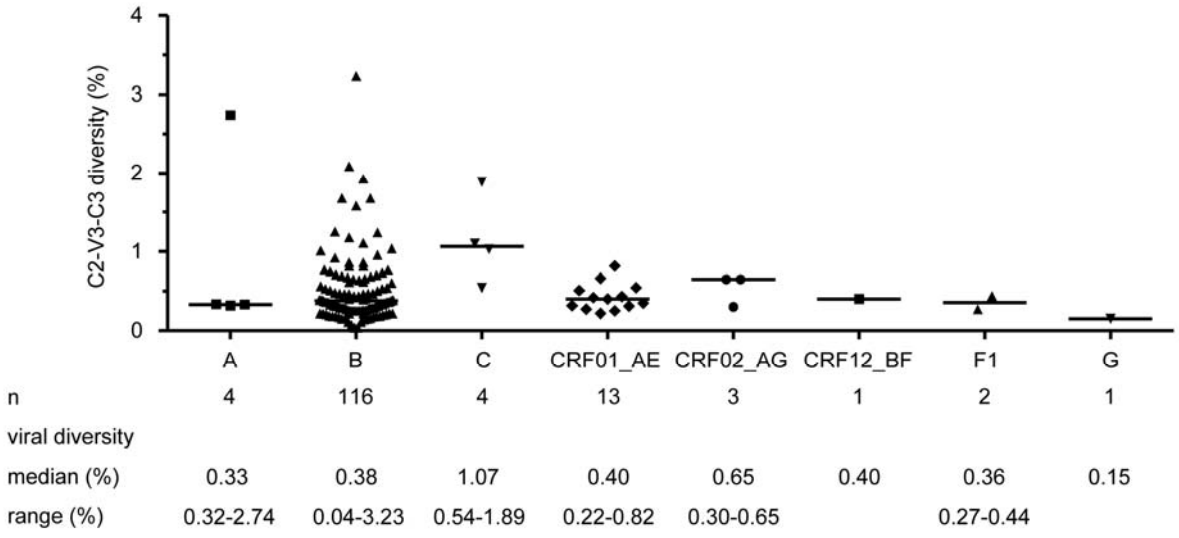

Figure 4. Influence of HIV-1 subtypes on viral diversity during primary HIV-1 infection. Viral diversities observed in the C2-V3-C3 region of the envelope gene in 145 primary HIV-1 infected patients are shown grouped according to HIV-1 subtypes. Bars indicate medians; CRF, circulating recombinant form.

Characteristics of patients with high viral diversity

We identified 16 subjects (11%) with a viral diversity higher than 1% despite a short estimated duration of infection (median 7 weeks; range 4-14). Of the 14 males 3 had a concomitant STI. One of the heterosexual men was infected either by sexual contact or by sharing needles. For detailed patient characteristics see table 2. Overall, no patient's or transmission's characteristics were overrepresented in the group of patients with viral diversity above 1%. These patients were infected with HIV-1 subtype A (1/4, 25%), B (12/116, 10%), and C (3/4; 75%).

Phylogenetic analyses of the C2-V3-C3 region in the 16 patients with viral diversities >1% showed 3 different tree patterns. In four patients, a starlike diversification was seen whereas in 8 patients the viral clones appeared subdivided into 2 to 4 subclusters each with diversity below 1% and average genetic distances between these intrapatient subgroups of 2.14% (range 1.11-4.37%). In the remaining 4 patients a mixture of these 2 described patterns was observed including subgroups showing viral diversity of more than 1%.

Table 2. Baseline characteristics of patients with higher and lower viral diversity during primary HIV-1 infection

viral diversity after transmission	**> 1%**		**< 1%**		
	n	**%**	**n**	**%**	
number of patients	16	100	129	100	
male	14	88	117	91	
female	2	13	12	9	
HIV-1subtype					
subtype A	1	6	3	2	
subtype B	12	75	104	81	
subtype C	3	19	1	1	
CRF01_AE			13	10	
transmission mode					
homosexual	10	63	96	74	
heterosexual	3	19	29	22	
sexually transmitted infections (STIs) [a]	3	19	17	13	
negative or indeterminate Western Blot	9	56	78	60	
	median	**(min-max)**	**median**	**(min-max)**	**p-value**
estimated duration of infection (weeks)	7	(4 - 14)	6	(2 - 24)	0.127 [c]
viral load (log_{10} copies of HIV-1 RNA / ml plasma)	5.1	(3.6 - 6.6)	5.4	(2.4 - 7.6)	0.623 [c]
CD4+ T-cells (cells / µl blood)	483	(127 - 1295)	411	(87 - 1120)	0.643 [c]
age (years)	32	(19 - 60)	36	(20 - 70)	0.403 [c]

[a] concomitant sexually transmitted infections: syphilis and/or chlamydia and/or gonorrhea

[b] median (min-max)

[c] Mann Whitney test

Viral tropism

Using PSSM, SVM_{wetcat}, and geno2pheno[coreceptor] to predict viral tropism in clonal C2-V3-C3 sequences, X4-tropic viruses or R5/X4 mixed populations were predicted for 10%, 19%, and 14% of the patients, respectively (table 3). In 102 patients (70%), all three prediction algorithms were consistent in predicting R5-tropic viral variants. The three bioinformatic tools yielded conflicting co-receptor usage predictions in 27% of the patients. In 4 patients (3%) all three tools predicted X4-tropic strains, either as a mixed population in three cases or as a pure X4-tropic population in one case. Full length *env* sequences obtained by SGA showed the same quasispecies composition and mutation patterns as clonal C2-V3-C3 sequences in these 4 patients (data not shown). Next, HIV-1 co-receptor usage was determined by phenotypic assays using 175 available primary isolates (from 117 patients) derived from plasma and PBMCs (Table 4). In a first step, all isolates were screened by the MT-2 cell assay and if X4-tropism was detected, this was confirmed by the GHOST cell assay allowing also differentiation of mixtures respectively of R5/X4 double users. Only in one of the 4 patients with concordant genotypic X4-tropism prediction and in one patient with concordant CCR5-tropism prediction a mixture of R5/X4-tropic viruses was detected in the PBMC isolate but not in the isolate obtained from plasma.

Table 3. Genotypic prediction of HIV-1 tropism in patients with primary HIV-1 infection

HIV-1 subtypes	all		B	
	n	%	n	%
number of patients	145	100	114	100
X4 tropism prediction by [a]				
Geno2pheno[coreceptor] [b]	21	14	13	11
PSSM [c]	15	10	5	4
SVM_{wetcat}	28	19	24	21
1 of 3 tools predicts X4 tropism	26	18	21	18
2 of 3 tools predict X4 tropism	13	9	6	5
3 of 3 tools predict X4 tropism	4	3	3	3
R5 tropism prediction by 3 tools	102	70	84	74

[a] X4 or R5/X4 mixed populations

[b] cutoff 5.75 (Prosperi *et al.* Retrovirology 2010, 7:56)

[c] using sinsi matrix

Table 4. Phenotypic prediction of HIV-1 tropism of primary HIV-1 isolates obtained from plasma and PBMCs of primary HIV-1 infected patients

genotype [a]		**phenotype**			
	number of patients predicted by 3 algorithms	number of patients analysed by MT-2 assay	X4	R5	R5/X4
X4	4	4	0	3 (75%)	1 (25%) [b]
R5	102	83	0	82 (98.8%)	1 (1.2%) [c]
inconsistent	39	30	0	30 (100%)	0

[a] Predicted by 3 bioinformatic tools: Geno2pheno, PSSM, SVM_{wetcat}

[b] Detected in PBMCs with MT-2 and GHOST cell assay

[c] Detected in PBMCs with MT-2 but not with GHOST cell assay

DISCUSSION

To investigate the complexity of transmitted HIV-1 populations, we determined viral diversity by HIV-1 envelope C2-V3-C3 clonal sequencing and assessed potential factors associated with transmission of heterogeneous viral populations in 145 patients enrolled in a single center primary HIV-1 infection cohort (ZPHI-study). This analysis revealed three major findings: 1) heterogeneous virus populations (diversity >1%) were found in 11% of all patients. 2) neither concomitant STI, gender, nor differences in sexual practices could be identified as factors associated with transmission of heterogeneous virus populations. 3) genotypic prediction of co-receptor tropism by three computational tools unambiguously characterized transmitted viruses of 73% of all patients: 70% of patients harboured R5-tropic viruses, and 3% (4 patients) were predicted as harbouring X4-tropic, respectively R5/X4-tropic viruses. However, in only one of those, CXCR4-usage could be confirmed phenotypically. In addition, one patient consistently predicted as harbouring R5-tropic viruses showed phenotypically a mixture of R5/X4-tropic viruses. Furthermore, results remained genotypically ambiguous in 27%.

Despite various efforts [2-5, 31, 32], factors associated with increased viral heterogeneity which is found in approximately 10-20% of PHI-patients have not been fully elucidated yet. However, this is not surprising given the difficulties to identify large populations of well characterized patients with PHI and the multi-factorial nature affecting diversity upon transmission. In particular, interpretation of results may be complicated by differences in patient populations with regard to time of sampling after infection, by disparities in ethnicity and living environments such as access to health care, sanitation conditions, and nutritional status, by potential selection bias of

patients enrolled in PHI studies, by uncertainties concerning sexual practices adopted, by the lack of knowledge of potential host genetic factors, and finally, by different methodologies used to investigate early transmitted viruses.

Our finding of complex HIV-1 populations in 11% of patients is in line with previous findings [2], although at a lower range. A likely explanation for these differences might be different frequencies of index patients transmitting a homogeneous virus population while they are in the acute phase of their HIV-1 infection. Another potential explanation is our rather conservative approach when compared to others [1] to search for complex viral populations using a cut-off of 1% diversity to differentiate homogeneous from heterogeneous transmitted viruses.

In contrast to previous studies [3, 31-33], no elevated complexity of transmitted viruses was found in the 20 patients with a concomitant STI during PHI. STIs clearly increase the risk of HIV-1 transmission by enhancing infectiousness and susceptibility [34]. The lower viral diversity in our patients might be due to the time of the transmission of the STI. Ongoing and already established STIs prior to HIV-1 infection may be of higher importance to render the mucosal barrier more susceptible for HIV-1 infection than those transmitted concurrently with HIV-1 or shortly thereafter. In our setting, the vast majority of patients reported first occurrence of STI symptoms very closely to ARS symptoms suggesting that STIs were co-transmitted with HIV-1. In contrast, Haaland *et al.* reported in their Zambian and Rwandan seroconverter patients that symptoms and signs of STIs were already present at the time of the last seronegative visit, suggesting that those STIs were present for comparatively longer durations, and therefore, mucosa may have been damaged more severely at the time of HIV-1 infection [31]. Another difference between cohorts

may be the proportion of female sex workers included [4]. In this group, the prevalence of STIs and the occurrence of mucosal microtraumata is higher than in males or females in general [3, 4, 35]. In our study, no female sex worker was participating.

HIV-1 subtype C may be associated with greater propensity for transmission possibility than other HIV-1 group M subtypes [34]. In our study the 4 patients infected with HIV-1 subtype C showed a higher median viral diversity compared to HIV-1 subtype B infected patients (p-value = 0.011). Viral diversity of CRF01_AE and other circulating recombinant forms were not significantly lower compared to HIV-1 subtype B, but none of those patients was infected with a more complex population (>1%). Whether higher diversity seen in our HIV-1 subtype C cases is based on HIV-1 subtype specific viral properties or not cannot be proven at this time. Similar diversities found in HIV-1 subtypes B and C in other PHI cohorts question the biological relevance of our finding [33]. However, it has to be noted that comparisons between those two HIV-1 subtypes were based on different study settings in different countries and different centers [33]. Thus, ultimately, only studies of larger patient cohorts enrolled under similar conditions will yield an answer to this question.

By combining three different genotypic prediction algorithms we found that 4 PHI-patients (3%) possibly harboured CXCR4-tropic viruses during acute HIV-1 infection. However, in only one of those patients, X4-tropic strains could be confirmed phenotypically. In addition, from one patient with consistent R5-coreceptor prediction X4-tropic strains could be isolated within a R5/X4 mixture. Thus, in our study transmission of relevant replication competent X4 variants seemed to be a rare event. These findings are in line with Raymond *et al.* [8] who reported 6.4% of

plasma viruses as X4-tropic respectively R5/X4-tropic mixtures in acutly HIV-1 infected patients. In contrast, a recent study using ultra-deep sequencing and genotypic prediction reported X4-tropic variants in approximately 50% of a small number of acutly HIV-1 infected patients studied [9]. Discrepant results are not surprising given the methodological difficulties regarding assays, prediction algorithms used, studied populations, and sample sizes. At present, the clinical relevance of transmitted X4-tropic minority variants is not known, because thresholds for frequencies of X4-tropic minority variants to predict maraviroc response during PHI are not available [36]. Furthermore, the problem with genotypic prediction is that it cannot predict replication competence of these minority species. On the other hand, phenotypic assays using primary viral isolates overcome this limitation but may lack sensitivity due to initial selection process during their propagation in cell cultures [37]. Recombinant phenotypic assays using PCR amplified HIV-1 envelope may have an improved sensitivity profile, but they give only information on tropism of single envelopes amplified and not on replication competence of these viruses. Taken together, despite a considerable amount of previous work [8-10, 38, 39], frequency of clinically relevant transmitted CXCR4-using viruses is still not known, but in our study transmission of such viruses seemed to occurr only rarely.

In summary, our findings suggest that transmission of complex virus founder populations may not depend solely on mucosal factors. In addition, transmission of clinically relevant CXCR4-using virus strains seems to remain a rare event and caution is warranted predicting co-receptor usage by genotypic algorithms alone, because concordance of those tools is still limited.

Acknowledgements

We are grateful to all patients participating in the ZPHI Study, Barbara Hasse, Urs Karrer, Rolf Oberholzer, Elisabeth Presterl, Reto Laffer, Klara Thierfelder, Yvonne Flammer, Markus Flepp and Thomas Frey for their dedicated patient care, Friederike Burgener and Dominique Klimpel for excellent laboratory assistance, Christine Vögtli and Ingrid Nievergelt for administrative support, and Alexander Thielen (Max Planck Institute for Informatics, Saarbrücken, Germany) for his help in analysing all our env sequences in parallel using geno2pheno[coreceptor]. We also thank the UZH IT services for giving us access to the Schrödinger high performance computer of the University of Zürich and Joseph K. Wong for helpful discussions and critical reading of the manuscript. We also thank and commemorate our dear friend and research colleague Marek Fischer for all his contributions and help to this project over many years. Very sadly, he died in December 2010 after a long illness.

References

1. Salazar-Gonzalez JF, Bailes E, Pham KT, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol **2008** Apr;82(8):3952-70.
2. Li H, Bar KJ, Wang S, et al. High multiplicity infection by HIV-1 in men who have sex with men. PLoS Pathog **2010** May;6(5):e1000890.
3. Sagar M, Lavreys L, Baeten JM, et al. Identification of modifiable factors that affect the genetic diversity of the transmitted HIV-1 population. Aids **2004** Mar 5;18(4):615-9.
4. Long EM, Martin HL, Jr., Kreiss JK, et al. Gender differences in HIV-1 diversity at time of infection. Nat Med **2000** Jan;6(1):71-5.
5. Bar KJ, Li H, Chamberland A, et al. Wide variation in the multiplicity of HIV-1 infection among injection drug users. J Virol **2010** Jun;84(12):6241-7.
6. Zhu T, Mo H, Wang N, et al. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. Science **1993** Aug 27;261(5125):1179-81.
7. Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. J Exp Med **1997** Feb 17;185(4):621-8.
8. Raymond S, Delobel P, Mavigner M, et al. CXCR4-using viruses in plasma and peripheral blood mononuclear cells during primary HIV-1 infection and impact on disease progression. AIDS **2010** Sep 24;24(15):2305-12.
9. Abbate I, Vlassi C, Rozera G, et al. Detection of quasispecies variants predicted to use CXCR4 by ultra-deep pyrosequencing during early HIV infection. **2011** Mar 13;25(5):611-7.
10. Curlin ME, Zioni R, Hawes SE, et al. HIV-1 envelope subregion length variation during disease progression. PLoS Pathog **2010**;6(12):e1001228.
11. Fatkenheuer G, Nelson M, Lazzarin A, et al. Subgroup analyses of maraviroc in previously treated R5 HIV-1 infection. N Engl J Med **2008** Oct 2;359(14):1442-55.
12. Gulick RM, Lalezari J, Goodrich J, et al. Maraviroc for previously treated patients with R5 HIV-1 infection. N Engl J Med **2008** Oct 2;359(14):1429-41.
13. Aceto L, Karrer U, Grube C, et al. [Primary HIV-1 infection in Zurich: 2002-2004]. Praxis (Bern 1994) **2005** Aug 10;94(32):1199-205.
14. Metzner KJ, Rauch P, von Wyl V, et al. Efficient suppression of minority drug-resistant HIV type 1 (HIV-1) variants present at primary HIV-1 infection by ritonavir-boosted protease inhibitor-containing antiretroviral therapy. J Infect Dis **2010** Apr 1;201(7):1063-71.
15. Rieder P, Joos B, von Wyl V, et al. HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men. AIDS **2010** May 15;24(8):1177-83.
16. Huber M, Fischer M, Misselwitz B, et al. Complement lysis activity in autologous plasma is associated with lower viral loads during the acute phase of HIV-1 infection. PLoS Med **2006**;3(11):e441.
17. Schmid A, Gianella S, von Wyl V, et al. Profound depletion of HIV-1 transcription in patients initiating antiretroviral therapy during acute infection. PLoS ONE **2010**;5(10):e13310.

18. Gianella S, vonWyl V, Fischer M, et al. Impact of early antiretroviral therapy during primary HIV-1 infection on cell-associated HIV-1 DNA and plasma HIV-1 RNA. Antiviral Therapy **2011**;in press.
19. Fischer M, Huber W, Kallivroussis A, et al. Highly sensitive methods for quantitation of human immunodeficiency virus type 1 RNA from plasma, cells, and tissues. J Clin Microbiol **1999** May;37(5):1260-4.
20. Joos B, Trkola A, Fischer M, et al. Low human immunodeficiency virus envelope diversity correlates with low in vitro replication capacity and predicts spontaneous control of plasma viremia after treatment interruptions. J Virol **2005** Jul;79(14):9026-37.
21. Sing T, Low AJ, Beerenwinkel N, et al. Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. Antivir Ther **2007**;12(7):1097-106.
22. Harada S, Koyanagi Y, Yamamoto N. Infection of HTLV-III/LAV in HTLV-I-carrying cells MT-2 and MT-4 and application in a plaque assay. Science **1985** Aug 9;229(4713):563-6.
23. Morner A, Bjorndal A, Albert J, et al. Primary human immunodeficiency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage. J Virol **1999** Mar;73(3):2343-9.
24. Japour AJ, Fiscus SA, Arduino JM, Mayers DL, Reichelderfer PS, Kuritzkes DR. Standardized microtiter assay for determination of syncytium-inducing phenotypes of clinical human immunodeficiency virus type 1 isolates. J Clin Microbiol **1994** Sep;32(9):2291-4.
25. Trkola A, Ketas T, Kewalramani VN, et al. Neutralization sensitivity of human immunodeficiency virus type 1 primary isolates to antibodies and CD4-based reagents is independent of coreceptor usage. J Virol **1998** Mar;72(3):1876-85.
26. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform **2008** Jul;9(4):286-98.
27. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **2002** Jul 15;30(14):3059-66.
28. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **2007** Aug;24(8):1596-9.
29. von Wyl V, Yerly S, Boni J, et al. Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. Arch Intern Med **2007** Sep 10;167(16):1782-90.
30. Fiebig EW, Wright DJ, Rawal BD, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS **2003** Sep 5;17(13):1871-9.
31. Haaland RE, Hawkins PA, Salazar-Gonzalez J, et al. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PLoS Pathog **2009** Jan;5(1):e1000274.
32. Sagar M, Kirkegaard E, Long EM, et al. Human immunodeficiency virus type 1 (HIV-1) diversity at time of infection is not restricted to certain risk groups or specific HIV-1 subtypes. J Virol **2004** Jul;78(13):7279-83.
33. Abrahams MR, Anderson JA, Giorgi EE, et al. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. J Virol **2009** Apr;83(8):3556-67.

34. Arien KK, Abraha A, Quinones-Mateu ME, Kestens L, Vanham G, Arts EJ. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. J Virol **2005** Jul;79(14):8979-90.
35. Poss M, Martin HL, Kreiss JK, et al. Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. J Virol **1995** Dec;69(12):8118-22.
36. Whitcomb JM, Huang W, Fransen S, et al. Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. Antimicrob Agents Chemother **2007** Feb;51(2):566-75.
37. Koot M, Vos AH, Keet RP, et al. HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay. AIDS **1992** Jan;6(1):49-54.
38. Frange P, Galimand J, Goujard C, et al. High frequency of X4/DM-tropic viruses in PBMC samples from patients with primary HIV-1 subtype-B infection in 1996-2007: the French ANRS CO06 PRIMO Cohort Study. J Antimicrob Chemother **2009** Jul;64(1):135-41.
39. Frange P, Chaix ML, Raymond S, et al. Low frequency of CXCR4-using viruses in patients at the time of primary non-subtype-B HIV-1 infection. J Clin Microbiol **2010** Oct;48(10):3487-91.

CHAPTER 2

HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men

Published as Rieder *et al*., AIDS 2010, 24:1177–1183

Own contribution

I participated in study conception and design. I participated in data acquisition, more precisely in *env* C2-V3-C3 clonal sequencing and full length *env* single genome sequencing, and analysis and interpretation of the sequence traces by bioinformatic tools. I did also the phylogenetic analysis using different bioinformatic tools. Further, I participated in acquisition, interpretation and analysis of clinical data. I also did the analysis and interpretation of the data. I prepared the manuscript and was involved in review and revision of the manuscript and approval prior to submission.

HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men

Philip Rieder[a,M], Beda Joos[a,M], Viktor von Wyl[a], Herbert Kuster[a], Christina Grube[a], Christine Leemann[a], Jürg Böni[b], Sabine Yerly[c], Thomas Klimkait[d], Philipp Bürgisser[e], Rainer Weber[a], Marek Fischer[a], Huldrych F. Günthard[a] and the Swiss HIV Cohort Study

[M] Philip Rieder and Beda Joos contributed equally to the writing of this article

[a]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich,
[b]Swiss National Center for Retroviruses, University of Zurich, Zurich,
[c]Laboratory of Virology, Division of Infectious Diseases, University Hospital Geneva, Geneva, [d]Institute of Medical Microbiology, University of Basel, Basel, and
[e]Division of Immunology, University Hospital Lausanne, Lausanne, Switzerland.

Abstract

Objective: To study transmission dynamics during acute infection, during the aviremic phase over the period of early antiretroviral therapy (ART) and during the phase of viral rebound after early treatment was stopped.

Methods: Transmission dynamics was assessed within 111 patients, enrolled in the Zurich primary HIV infection study, by molecular epidemiological methods using *pol* sequences from genotypic resistance tests and clonal *env* C2–V3–C3 sequences. Coclustering of Zurich primary HIV infection sequences with 12 303 sequences from 8837 HIV-positive patients enrolled in the multisite Swiss HIV Cohort Study was identified. Furthermore, we investigated transmission patterns within phylogenetic clusters by using longitudinal clinical data and analyzed HIV transmission by stage of infection and attempted to localize transmission events to periods before or after early ART.

Results: Six transmission clusters comprising 20 men having sex with men were identified. Furthermore, linkage to eight men having sex with men from the Swiss HIV Cohort Study could be established. Strikingly, we detected at least five new primary infection events originating from Zurich primary HIV infection patients within 16–61 weeks after stopping early ART. Viral loads of likely index patients varied from 314 up to 1 690 000 HIV-1 RNA copies/ml of plasma at the estimated time of infection.

Conclusion: The large number of new infections originating from men having sex with men who stopped early ART indicates that current preventive efforts are insufficient. In contrast, these patients showed no adherence problems. These findings argue for early, continuous ART in sexually active HIV-1-infected persons not only for individual patient benefits but also specifically to reduce the spread of HIV-1.

Introduction

HIV-associated morbidity and mortality have declined among patients with access to antiretroviral therapy (ART) [1]. However, this success was not matched by similar reductions in HIV transmission. Current data even suggest a recent increase in HIV diagnoses in resource-rich settings, particularly among men having sex with men (MSM) [2–4]. Both acute and late stages of HIV-1 infection are associated with high viral loads and, therefore, may disproportionately contribute to the spread of the epidemic [5–7]. In-depth knowledge of the dynamics of HIV-1 transmission within specific transmission groups is fundamental to develop effective prevention strategies [8,9]. Phylogenetic analysis represents one strategy for investigating transmission dynamics between individuals, and occasionally also for forensic use [10–13]. Recent studies have used pol sequences generated for drug resistance genotyping [5,14–16] to reconstruct transmission histories [17].

Here, we analyzed transmission networks within 111 well documented patients enrolled in the Zurich primary HIV infection (ZPHI) study [18–20] by molecular epidemiological data based on two HIV genes. Furthermore, we studied viral genetic linkage with persons enrolled in the multicenter Swiss HIV Cohort Study (SHCS), which includes data on approximately 50% of all HIV-infected and 75% of all treated individuals in the country and thus reaches an otherwise unmatched representativity [15,21,22]. For the linkage studies, we used *pol* sequences from the Swiss HIV drug resistance database with additional confirmation through sequencing of clonal HIV *env* C2–V3–C3 fragments. In particular, we sought to identify transmission events in the viremic phase during acute infection, the aviremic phase during early ART and during the phase of viral rebound after early treatment was stopped.

Methods

Our analysis included clinical and genotypic data from the SHCS (for detailed description, see Supplemental Digital Content, http://links.lww.com/QAD/A40), a clinic-based study with continuous enrolment and at least semiannual visits [23,24]. All available HIV pol sequences (12 303) obtained from genotypic resistance tests were used for phylogenetic analysis to select candidates for confirmatory env sequencing.

Patients enrolled in the ZPHI study (http://clinicaltrials. gov: NCT00537966) [18–20,25] presented with documented acute or recent primary HIV-1 infection (for definitions and detailed patient characteristics, see Supplement, http://links.lww.com/QAD/A40). They were offered early ART and could chose after 1 year successful treatment (viral load <50 copies/ml plasma) to stop it. Blood samples were collected at the time of enrolment and sequentially in at least 3-month intervals. The studies complied with the principles of the Declaration of Helsinki and the guidelines of the local ethical committee. Written informed consent was obtained from all participants. Patient data were anonymized for all analyses.

For each patient, a minimal and maximal timepoint of infection was estimated integrating all available information, including patient history relating to known risk situations, occurrence of first symptoms, previous negative test results, avidity assays and western blot. This conjectured time after infection (mean of upper and lower-bound estimates) was positively correlated with env diversity ($r^2 = 0.07$, $P = 0.05$). Transmission dates were also used to calculate cluster-specific infection rates (for detailed algorithms, see Supplement, http://links.lww. com/QAD/A40).

RNA extraction, amplification, cloning and sequencing of HIV-1 *env* C2–V3–C3 fragments was performed by modification of previously described methods [26,27] (for detailed description, see Supplement, http://links. lww.com/QAD/A40). Sequences were edited and aligned with SeqMan (DNASTAR Inc., Madison, Wisconsin, USA). Alignments were refined with MAFFT [28] and manually adjusted. Genetic distance estimates were obtained by Molecular Evolutionary Genetics Analysis software [29] using the Tamura–Nei model. Neighbor-joining phylogenetic trees were constructed using HXB2 and Non-B strains as references and bootstrapping (1000 replications). Maximum-likelihood trees were inferred by the DNA Maximum Likelihood program using randomized input order, global

rearrangements and multiple jumble options (PHYLIP3.6, distributed by J. Felsenstein). The reported *env* sequences were deposited in GenBank under accession numbers GU471280–GU471687.

Statistical analyses were performed using Prism5 (Graph-Pad Software, San Diego, California, USA) and Stata10.1 (Stata Corp., College Station, Texas, USA). Nonparametric tests (Mann–Whitney) were used for group comparison.

Results

We analyzed patients enrolled in the ZPHI study before December 2007. One hundred eleven patients were followed over 3.3 (median, range 0.5–6.7) years (for details see, Table S1, http://links.lww.com/QAD/A40). Ninety-three patients underwent early ART that was stopped in 51% of them (n = 47) after approximately 1 year of viral suppression. Baseline blood samples were available within 5 (2–13) weeks after the estimated timepoint of infection in acute and within 13.5 (7–24) weeks in recently infected patients. To investigate transmission dynamics within the ZPHI study and to identify transmission networks 6500 *pol* sequences from 4276 well characterized SHCS patients were used along with those from the ZPHI patients to generate a neighbor-joining phylogenetic tree. Clusters, containing at least two ZPHI patients, with bootstrap values above 98% and genetic distances below 1.5% (consistent with [5,17]) were preselected. Then, in an independent analysis exploiting the genetic information inherent to the highly variable envelope V3 region, we confirmed the suggested transmission clusters by constructing maximum likelihood trees with clonal *env* sequences of the 111 ZPHI patients as well as from the SHCS patients clustering to ZPHI patients within the *pol* phylogeny.

Six clusters containing 20 ZPHI patients (18%) and eight SHCS patients were identified (Fig. 1). No dual infections and no transmitted major drug resistance mutations [30] were observed in these populations. Transmission of minor populations carrying resistance mutations M184V and K103N have been observed in 13.9% of patients in the ZPHI study [21]. The individual transmission events were studied among these patients within a median follow-up time of 4.0 (0.5–7.6) years (Fig. 2, Table S2, http:// links.lww.com/QAD/A40). The extent of observed transmissions may be underestimated, as at the time of analysis some potential transmitters have not yet been diagnosed. Therefore, we performed a second phylogenetic analysis in May 2009 including 12 300 pol sequences of the growing SHCS dataset from 8837 patients.

In cluster A, ZPHI-A1 and ZPHI-A2 formed a serodiscordant couple living in monogamous relationship. ZPHI-A2 has been infected 304 days after ZPHI-A1 stopped continuing early ART. Transmission occurred during the chronic phase of the index partner while his viral load was 33 478 copies/ml.

In cluster B, ZPHI-B2 has been infected most likely by passive anal intercourse

during the chronic phase of ZPHI-B1 136 days after discontinuing early ART. The viral load at the estimated timepoint of infection was 930 copies/ml.

ZPHI-C1 and ZPHI-C2 together consulted our outpatient clinic because they suspected HIV infection of ZPHI-C2 due to symptoms of an acute retroviral syndrome. Actually, ZPHI-C1 infected ZPHI-C2 146 days after stopping ART (viral load 2237 copies/ml) during chronic phase. Because they formed a monogamous relationship at that time the two additional SHCS patients within this cluster could be excluded as potential transmitters to patient ZPHI-C2. ZPHI-C1 was infected approximately 2 years earlier, possibly by SHCS-C4.

ZPHI-D1 appeared phylogenetically linked to ZPHI-D2 (genetic distance 1.26% in pol, 0.92% in *env*). However, ZPHI-D1 was aviremic (viral load <50 copies/ml) at the timepoint of infection of ZPHI-D2 due to successful ART. This transmission pair, therefore, seems unlikely, as proven transmissions with undetectable plasma viremia have never been reported [31]. Moreover, in the extended second analysis, three other potential transmitters were found.

In cluster E, ZPHI-E1 may have infected ZPHI-E2 8 days after having been infected. The viral load at this potential transmission event was 1 690 000 copies/ml. Both, E1 or E2 may have infected ZPHI-E3 during their recent phase (mean 132 days after infection; max viral load 1039 copies/ml). ZPHI-E4 and ZPHI-E5 likely have been infected by either ZPHI-E2 or ZPHI-E3 after stopping early ART but not by ZPHI-E1 who continued suppressive therapy (Table S2, http://links.lww.com/ QAD/A40).

Although the complexity in cluster F with six SHCS patients and seven ZPHI patients prohibited analysis of the transmission dynamics, sources of new infections seem to be untreated chronically infected rather than acutely infected patients. Chronically infected patients SHCS-F8, F9 and F10 did not receive treatment during several years, when ZPHI-F1, ZPHI-F3, ZPHI-F6, ZPHI-F4, ZPHI-F2, ZPHI-F5 and ZPHI-F7 were infected sequentially. All of the latter ones reached undetectable viral load during most of the timepoints when those new infection events happened because they received early ART.

Taken together, these patients demonstrate the importance of using clinical, laboratory and epidemiological data to supplement phylogenetic analysis in the assessment of putative transmission chains [32]. Surprisingly, only in one example transmission may have happened during the acute phase (ZPHI-E1, ZPHI-E2) and in one patient within the recent phase (ZPHI-E3) of the possible source. However, in

five other patients, transmission presumably took place while the index patient was already in the chronic stage of infection 109– 425 days after interruption of early ART. Notably, under treatment, all patients showed undetectable viral load (except few blips) indicating that adherence to therapy was generally very good. Taking into account the overall time when patients were under virologically suppressive treatment, we estimate that 3.5 [95% confidence interval (CI) 0.9–13.5] infection events per person-year occurred prior to treatment initiation and 1.8 (95% CI 0.5–5.8) events per person-year after cessation of the initial treatment.

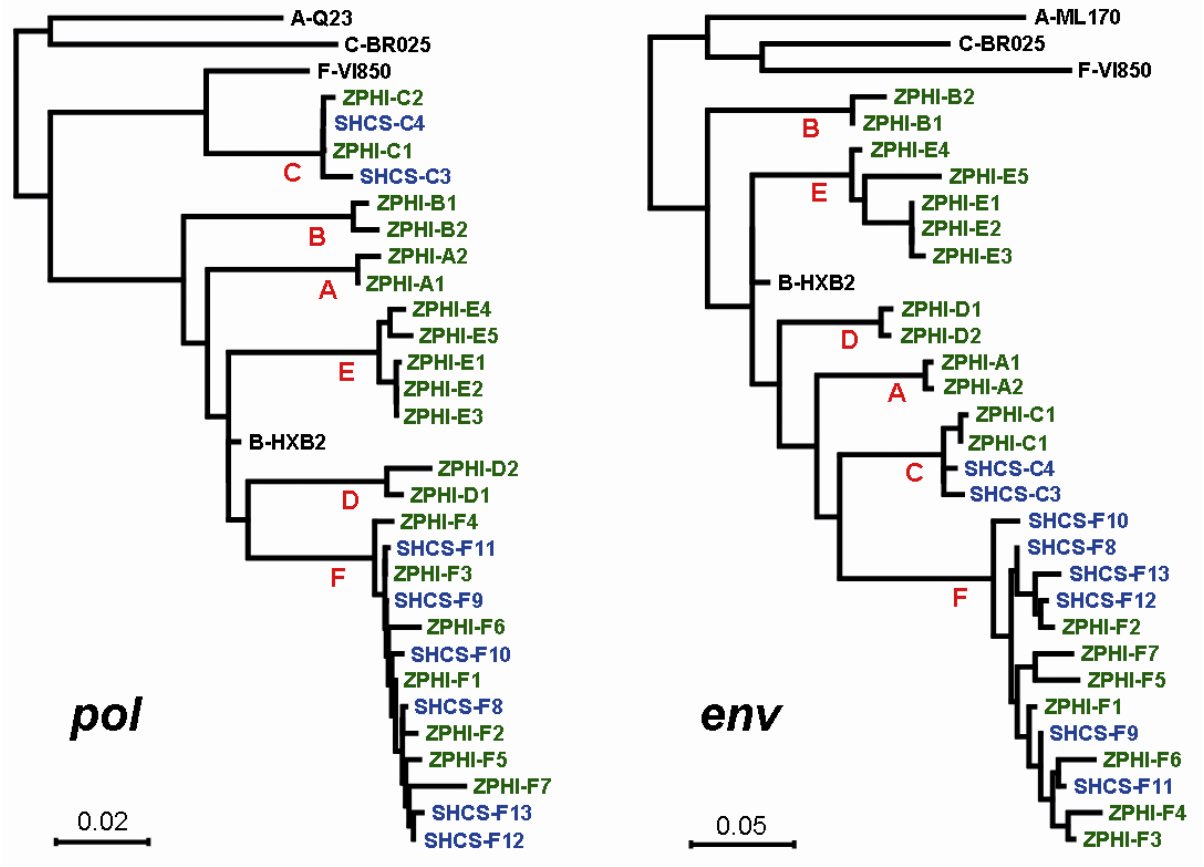

Fig. 1. Transmission clusters inferred from *pol* and *env* sequences. Maximum likelihood phylogenetic trees of *pol* sequences (left panel) and consensus sequences of clonal HIV-1 *env* C2–V3–C3 sequences (right panel) isolated from ZPHI patients (green) and SHCS patients (blue). Broadly similar trees were inferred by neighbor-joining method. Transmission groups (A–F) formed six robust clusters (bootstrap support 99–100%). Reference strains of different HIV subtypes were included as outgroup. SHCS, Swiss HIV Cohort Study; ZPHI, Zurich primary HIV infection study.

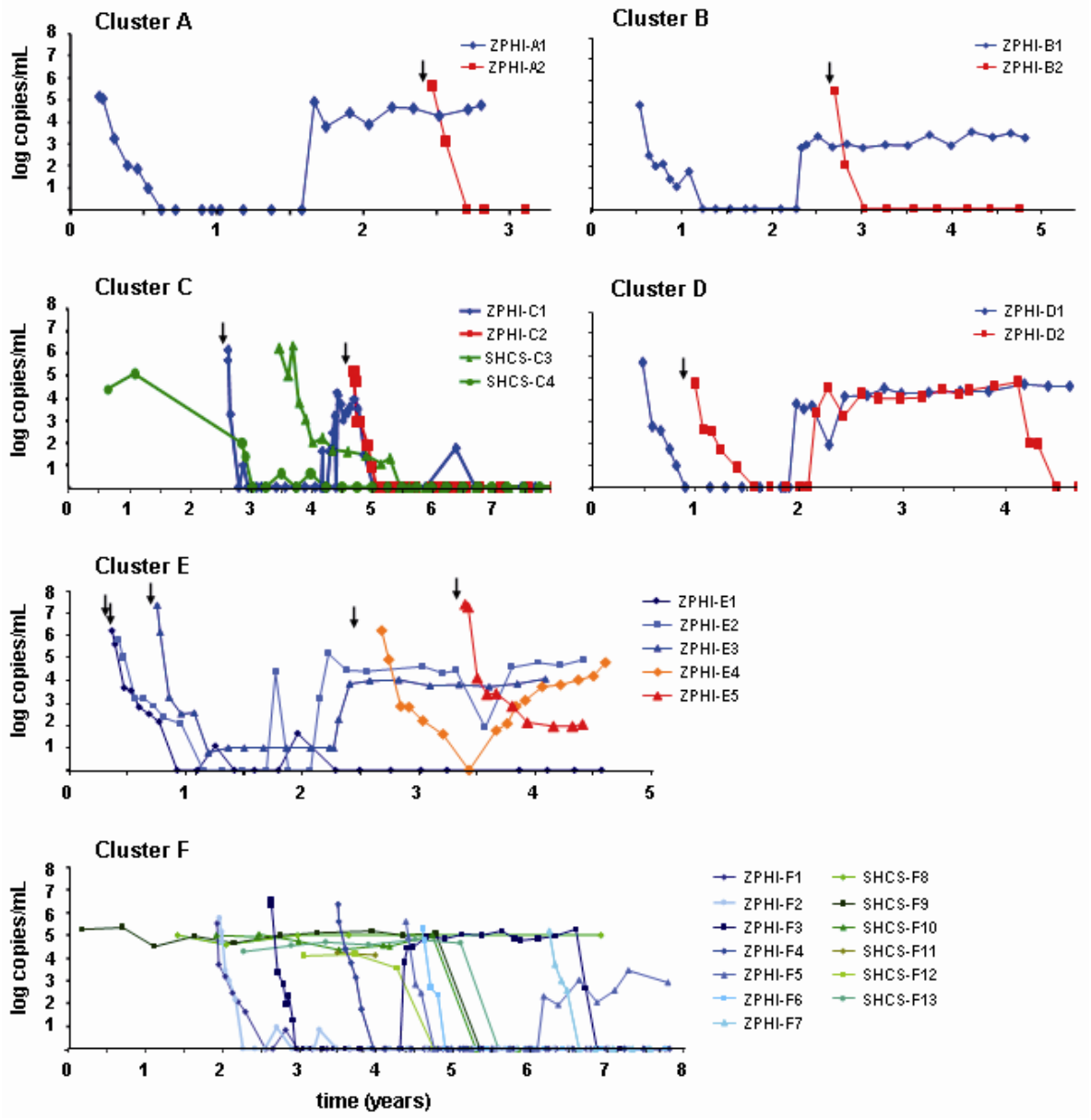

Fig. 2. Time course analysis of transmission clusters. Time course of viral loads of HIV-1-infected patients within clusters A to F (assumed transmission chains). In cluster A ZPHI-A2 was infected 304 days after early ART stop of the possible transmitter ZPHI-A1. In clusters B and C, the infection occurred in the same way only 136 and 146 days, respectively, after treatment stop. In addition, ZPHI-C1, ZPHI-C2 and ZPHI-A1, ZPHI-A2 have been confirmed as couples. In the phylogenetic cluster D, ZPHI-D1 would have infected ZPHI-D2 with a viral load of 50 copies/ml indicating that a third party was involved in this transmission chain. In the transmission chain of cluster E, ZPHI-E4 was infected 82–136 days and ZPHI-E5 398–452 days after stop of early ART of ZPHI-E2 and ZPHI-E3. Because of the complexity of cluster F, index partners cannot be identified but are most probably chronically infected patients enrolled in the SHCS. Black arrows indicate estimated timepoints of infection. ART, antiretroviral therapy; SHCS, Swiss HIV Cohort Study; ZPHI, Zurich primary HIV infection study.

Discussion

In this study, we took advantage of our well characterized ZPHI cohort [18–20] and the sequence database linked to the clinical data of the SHCS [24]. In addition to phylogenetic analysis of pol sequences confirmed by sequencing of clonal env fragments, we estimated the timepoint of infection and followed 111 patients longitudinally over years. Eighteen percent of the ZPHI patients formed transmission clusters. We further dissected eight transmission events in five phylogenetic clusters; only one transmission occurred during acute and another one during the recent phase. However, five transmissions occurred during chronic stage of the presumed transmitters, more than 3 months after stopping early ART. This was unexpected and worrisome because it shows the limitations of prevention measures in this sexually active MSM cohort. Infectiousness during chronic infection was quite high in this population also at relatively low viral load in some cases (range 314– 1 690 000 copies/ml).

In previous studies using phylogenetic reconstruction, patients were categorized as acutely or chronically infected according to their stage at diagnosis but not when transmission actually occurred [5,32]. This probably led to overestimation of transmission frequencies during acute/early infection. Phylogenetic analyses have limitations, as one can never rule out transmissions potentially originating from other index cases not known to the investigator. This effect is attenuated in our setting because the primary HIV infection patients enrolled represent approximately 55% of all newly HIV-infected MSM patients in the canton of Zurich and the likelihood that HIV-infected patients are enrolled in the SHCS is high [21,22]. In contrast to other phylogenetic studies assessing transmission dynamics [5,15,33,34], we used two different samples and analyzed two different genetic regions to increase genetic information and to exclude laboratory and database errors [13]. Moreover, plausibility of the transmission clusters was controlled by longitudinal viral load data.

Conclusion

This study demonstrates that in our intensely studied sexually active MSM, collective preventive safer-sex counselling was insufficient as documented by the high number of new infections originating from patients who stopped early ART according to study protocol. Of note, the same individuals were very adherent to ART. Furthermore, we detected a remarkable proportion of new infections originating from index patients being already in their chronic phase, sometimes with low viral loads. These findings argue strongly for early, continuous ART in sexually active HIV-1-infected MSM. This strategy, most likely, will have a profound impact to reduce the further spread of HIV-1.

Acknowledgements

This study has been financed in the framework of the SHCS, supported by the Swiss National Science Foundation (SNF #33CSC0-108787). Further support was provided by the SNF grant #3247B0-112594 and #320000-116035 (to H.F.G.), the Union Bank of Switzerland in the name of a donor to H.F.G., unrestricted research grant from Tibotec, Switzerland, the SHCS research foundation and SHCS projects #470 and #528.

We are grateful to all the patients participating in the ZPHI Study and in the SHCS; Barbara Hasse, Urs Karrer, Rolf Oberholzer, Elisabeth Presterl, Reto Laffer, Ulrich von Both, Klara Thierfelder, Dominique Braun, Markus Flepp and Thomas Frey for their dedicated patient care; Friederike Burgener and Dominique Klimpel for excellent laboratory assistance; Christine Vögtli and Ingrid Nievergelt for administrative support. We also would like to thank the UZH IT services for giving us access to the highperformance computer of the University of Zurich. Furthermore, we thank all the staff of the SHCS clinical centers, the data center and all resistance laboratories for their great work and Joseph K. Wong for critical review of the manuscript.

Günthard conceptualized, designed and supervised the study. Data acquisition was done by Rieder, Joos, von Wyl, Kuster, Grube, Leemann, Böni, Yerly, Klimkait, Bürgisser and Günthard. Analysis of the manuscript was done by Rieder, Joos, Günthard, and von Wyl. Critical revision of this article was done by Fischer, Weber, Kuster, Bürgisser, Yerly, Böni and Klimkait.

H.F.G. has been an adviser, consultant or both for the following companies: GlaxoSmithKline, Abbott, Novartis, Boehringer-Ingelheim, Roche, Tibotec and Bristol-Myers Squibb, and has received unrestricted research and educational grants from Roche, Abbott, Bristol-Myers Squibb, GlaxoSmithKline, Tibotec and Merck Sharp & Dohme. S.Y. has participated in advisory boards of Bristol-Myers Squibb and Tibotec, and has received travel grants from GlaxoSmithKline and Merck Sharp & Dohme. V.V.W., P.R., H.K., C.G., C.L., J.B., T.K., P.B., R.W. and M.F. have no conflicts of interest.

References

1. Egger M, Hirschel B, Francioli P, Sudre P, Wirz M, Flepp M, et al. Impact of new antiretroviral combination therapies in HIV infected patients in Switzerland: prospective multicentre study. Swiss HIV Cohort Study. BMJ 1997; 315:1194–1199.
2. BAG BfG. HIV/AIDS in Switzerland on September 30th 2008. (HIV/AIDS in der Scweiz am 30. September 2008.) In: Bulletin 44. Bern, Switzerland: Bundesamt fu¨r Gesundheit BAG; 2008. pp. 778–781.
3. Marcus U, Voss L, Kollan C, Hamouda O. HIV incidence increasing in MSM in Germany: factors influencing infection dynamics. Euro Surveill 2006; 11:157–160.
4. Fisher M, Pao D, Murphy G, Dean G, McElborough D, Homer G, Parry JV. Serological testing algorithm shows rising HIV incidence in a UK cohort of men who have sex with men: 10 years application. AIDS 2007; 21:2309–2314.
5. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, et al. High rates of forward transmission events after acute/ early HIV-1 infection. J Infect Dis 2007; 195:951–959.
6. Pilcher CD, Tien HC, Eron JJ Jr, Vernazza PL, Leu SY, Stewart PW, et al. Brief but efficient: acute HIV infection and the sexual transmission of HIV. J Infect Dis 2004; 189:1785–1792.
7. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med 2000; 342:921–929.
8. Salomon JA, Hogan DR, Stover J, Stanecki KA, Walker N, Ghys PD, Schwartlander B. Integrating HIV prevention and treatment: from slogans to impact. PLoS Med 2005; 2:e16.
9. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 transmission, by stage of infection. J Infect Dis 2008; 198:687–693.
10. Albert J, Wahlberg J, Leitner T, Escanilla D, Uhlen M. Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 pol and gag genes. J Virol 1994; 68:5918– 5924.
11. Birch CJ, McCaw RF, Bulach DM, Revill PA, Carter JT, Tomnay J, et al. Molecular analysis of human immunodeficiency virus strains associated with a case of criminal transmission of the virus. J Infect Dis 2000; 182:941–944.
12. Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, Korber BT, et al. Molecular epidemiology of HIV transmission in a dental practice. Science 1992; 256:1165–1171.
13. Leitner T, Albert J. Reconstruction of HIV-1 transmission chains for forensic purposes. AIDS Rev 2000; 2:241–251.
14. Hue S, Gifford RJ, Dunn D, Fernhill E, Pillay D. Demonstration of sustained drug-resistant human immunodeficiency virus type 1 lineages circulating among treatment-naive individuals. J Virol 2009; 83:2645–2654.
15. Yerly S, Junier T, Gayet-Ageron A, Amari EB, von Wyl V, Günthard HF, et al. The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. AIDS 2009; 23:1415–1423.
16. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med 2008; 5:e50.

17. Hue S, Clewley JP, Cane PA, Pillay D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. AIDS 2004; 18:719–728.
18. Aceto L, Karrer U, Grube C, Oberholzer R, Hasse B, Presterl E, et al. Primary HIV-1 infection in Zurich: 2002–2004. Praxis (Bern 1994) 2005; 94:1199–1205.
19. Huber M, von Wyl V, Ammann CG, Kuster H, Stiegler G, Katinger H, et al. Potent human immunodeficiency virusneutralizing and complement lysis activities of antibodies are not obligatorily linked. J Virol 2008; 82:3834–3842.
20. Rusert P, Kuster H, Joos B, Misselwitz B, Gujer C, Leemann C, et al. Virus isolates during acute and chronic human immunodeficiency virus type 1 infection show distinct patterns of sensitivity to entry inhibitors. J Virol 2005; 79:8454–8469.
21. Metzner KJ, Rauch P, von Wyl V, Leemann C, Grube C, Kuster H, et al. Efficient suppression of minority drug-resistant HIV type 1 (HIV-1) variants present at primary HIV-1 infection by ritonavir-boosted protease inhibitor-containing antiretroviral therapy. J Infect Dis 2010; 201 (7):1063–1071.
22. Schoeni-Affolter F, Rickenbach M, Furrer H, Rudin C, Günthardt HF, Ledergerber B, et al. Cohort Profile: The Swiss HIV Cohort Study. Int J Epidemiol, Advance Access published on November 30, 2009; doi:10.1093/ije/dyp321
23. Ledergerber B, Egger M, Opravil M, Telenti A, Hirschel B, Battegay M, et al. Clinical progression and virological failure on highly active antiretroviral therapy in HIV-1 patients: a prospective cohort study. Swiss HIV Cohort Study. Lancet 1999; 353:863–868.
24. von Wyl V, Yerly S, Boni J, Burgisser P, Klimkait T, Battegay M, et al. Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. Arch Intern Med 2007; 167:1782–1790.
25. Huber M, Fischer M, Misselwitz B, Manrique A, Kuster H, Niederöst B, et al. Complement lysis activity in autologous plasma is associated with lower viral loads during the acute phase of HIV-1 infection. PLoS Med 2006; 3:e441.
26. Fischer M, Huber W, Kallivroussis A, Ott P, Opravil M, Luthy R, et al. Highly sensitive methods for quantitation of human immunodeficiency virus type 1 RNA from plasma, cells, and tissues. J Clin Microbiol 1999; 37:1260–1264.
27. Joos B, Trkola A, Fischer M, Kuster H, Rusert P, Leemann C, et al. Low human immunodeficiency virus envelope diversity correlates with low in vitro replication capacity and predicts spontaneous control of plasma viremia after treatment interruptions. JVirol 2005; 79:9026–9037.
28. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002; 30:3059–3066.
29. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24:1596–1599.
30. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M, et al. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PLoS One 2009; 4:e4724.
31. Vernazza P, Hirschel B, Bernasconi E, Flepp. Seropositve per-sons without other STDs and receiving effective treatment cannot transmit the virus through sexual relations. (Les per sonnes seropositives ne souff rant d'aucune autre MST et suivant un traitment antiretroviral ef cie ne transmettent pas le VIH par voie sexuelle.). Bull Med Suisses 2008; 89:165–169.

32. Brown AE, Gifford RJ, Clewley JP, Kucherer C, Masquelier B, Porter K, et al. Phylogenetic reconstruction of transmission events from individuals with acute HIV infection: toward more-rigorous epidemiological definitions. J Infect Dis 2009; 199:427–431.
33. Pao D, Fisher M, Hue S, Dean G, Murphy G, Cane PA, et al. Transmission of HIV-1 during primary infection: relationship to sexual risk and sexually transmitted infections. AIDS 2005; 19:85–90.
34. Brenner BG, Roger M, Moisi DD, Oliveira M, Hardy I, Turgel R, et al. Transmission networks of drug resistance acquired in primary/early stage HIV infection. AIDS 2008; 22:2509–2515.

CHAPTER 3

Rational design of HIV-1 fluorescent hydrolysis probes considering phylogenetic variation and probe performance

Published as Althaus et al., J. Virol Methods. 2010 May;165(2):151-60

Own contribution

I contributed in design of FH-probes. I designed phylogenetically conserved FH-probes by hand as a prerequisite for automatization. I was also involved in review and revision of the manuscript.

Rational design of HIV-1 fluorescent hydrolysis probes considering phylogenetic variation and probe performance

Claudia F. Althaus*[a], Sara Gianella*[a], Philip Rieder*[a], Viktor von Wyl[a], Roger D. Kouyos[b], Barbara Niederöst[a], Adrian Schmid[a], Karin J. Metzner[a], Beda Joos[a], Huldrych F. Günthard[a] and Marek Fischer[a§]

* These authors contributed equally to this work,

[a] Division of Infectious Diseases, University Hospital Zurich, University of Zurich, Switzerland,

[b] Institute of Integrative Biology, ETH Zurich, Switzerland, [§] Corresponding author

Abstract

Quantitative PCR (qPCR) using fluorescent hydrolysis probes (FH-probes; TaqMan®-probes) of variable genomes, such as HIV-1, can result in underestimation of viral copy numbers due to mismatches in the FH-probe's target sequences. Therefore both target conservation and physical properties of FH-probes, such as melting temperature, baseline fluorescence and secondary structure, should be considered in design of FH-probes.

Analysis of a database of 1242 near full-length HIV-1 sequences with a novel computational tool revealed that the probability of target and FH-probe identity decreases exponentially with FH-probe length. In addition, this algorithm allowed for identification of continuous sequence stretches of high conservation, from which FH-probes with global HIV-1 clade coverage could be chosen. To revise the prerequisites of physical FH-probe function, properties of 30 DNA and 21 chimeric DNA locked nucleic acid (DLNA) HIV-1 FH-probes were correlated with their performance in qPCR. This identified the presence of stable secondary structures within FH-probes and the base composition and thermal stability of the 5'-proximal end as novel predictors of FH-probe performance.

Thus, empirically validated novel principles of FH-probe design regarding conservation and qPCR-performance were identified, which complement and extend current rules for FH-probe design.

Abbreviations

Ct, cycle threshold; Delta-Tm, Tm-difference of FH-probe and the primer binding the same strand; DLNA, chimeric DNA locked nucleid acid; FH-probe, fluorescent hydrolysis probe; qPCR, quantitative PCR; RT-qPCR, reverse transcriptase qPCR; SNR, signal to noise ratio; Tm, melting temperature

1. Introduction

Monitoring qPCR by FH-probes encounters its major pitfalls in assays examining phylogenetically diverse viruses such as HIV-1. While primers are relatively forgiving in respect to target variation[1], as little as one mismatch in a FH-probe's target sequence can greatly perturb the results of qPCR by underestimating copy numbers[2]. This feature of FH-probes is widely exploited in genotyping of single nucleotide polymorphisms[3,4].

Further factors influencing the function of FH-probes are related to their physical properties, which are determined by primary and secondary nucleic acid structure. Hybridization characteristics and kinetics of the FH-probe to its target have been assumed to be crucial for the performance of FH-probes. Hence, to ensure optimal binding of FH-probe to target, a choice of binding sites devoid of secondary structures and with 40-60% G/C content has been proposed[5-9]. In addition, to ensure hybridisation of the FH-probe before elongation of the newly synthesized DNA strand, it has been postulated that the melting temperature (Tm) of FH-probes should be 5-10°C higher than the Tm of the primer binding to the same DNA strand as the FH-probe[5,6].

Performance of FH-probes in qPCR assays also depends on intensity of fluorescence, which is suppressed in the absence of PCR target by a quenching moiety commonly placed at the 3' end of the FH-probe and elicited by nucleolytic detachment of the base carrying the fluorophore, usually placed at the FH-probe's 5' end. The physical proximity of fluorophore and quencher in the absence of PCR product has been reported to influence performance of FH-probes[9]. Optimal proximity of the two modified bases in the free FH-probe may be attained by keeping the FH-probe length minimal[10] or, if long FH-probes need to be designed, by attachment of the quencher to internal positions[11].

Another important rule advises to avoid guanosine at the residue carrying the fluorophore because it quenches fluorescence even after hydrolysis and it may be cleaved off with reduced efficiency[9,12].

In the present study, the phylogenetic constraints for design of FH-probes were addressed by analyzing the relationship of FH-probe match frequency with FH-probe length and single base conservation. Physical properties of FH-probes and their

influence on FH-probe performance were examined by testing 51 probes targeting 6 amplica (plural of amplicon) within the HIV-1 genome.

2. Methods

2.1. Conservation of FH-probes

The conservation of the chosen amplica was calculated by taking into account all circulating HIV-1 groups (subtypes M, N, O, CPZ). The sequences were obtained from curated alignments in the Los Alamos HIV Sequence Database (download 2007, www.hiv.lanl.gov) comprising a total of 1242 sequences including 17% subtype-B isolates. Due to location at the ends of the genome, the *early* and the *nef* amplica, respectively, were only covered by 233 (34% subtype B) and 1231 (16.5% subtype B) viral isolates. The levels of single base conservation were compared to a global consensus and calculated using Bioedit[13].

Identification of short continuous stretches of matching oligonucleotide sequences was carried out using the R statistical computing environment (version 2.8.1)[14] and tools from the seqinR package available at http://pbil.univ-lyon1.fr/software/SeqinR. Briefly, after removal of gaps, this implementation locates all possible substrings of a defined length within a reference sequence and calculates the corresponding frequencies of perfect matches to the alignment. Each amplicon was analyzed separately. The source code is available upon request.

Match frequencies (Table 1) of FH-probes were calculated based for either only B-subtypes or all HIV-1 groups (subtypes M, N, O, CPZ).

2.2. FH-probe design

All FH-probes contained the fluorophore FAM and the quencher TAMRA. Tms of the DNA FH-probes were calculated with an open source program (www.biophp.org) using the base-stacking algorithm, which was adjusted for the presence of 3mM $MgCl_2$. In all FH-probes a 5' terminal G was avoided[5,6,8]. The calculated Tms were aimed to exceed 55°C, except for mf341, which was designed solely based on phylogenetic considerations. DLNA FH-probes were designed with a length of 8-14 bases and with a content of LNA bases of 10-75%. Tm for DLNA FH-probes was calculated with an open source program different from that used for DNA FH-probes (http://lna-tm.com), considering differences in melting temperature between DNA and LNA bases.

2.3. Tm measurement

The Tm of the FH-probes was measured by examining the denaturation kinetics of the FH-probe (0.1uM) and its complementary strand (1uM) in a range from 30-90°C in 0.5°C intervals for 10 seconds using a real-time thermocycler (IQ5, Biorad, Basel, Switzerland). To attain conditions close to those of qPCR, the complementary strands were designed with overhangs of 3 bases avoiding dangling ends and the experiments were performed in 1 x Hotstartaq buffer (Qiagen) lacking enzyme with the inclusion of 3mM $MgCl_2$. The analysis of the melting curves was executed by subtracting the background fluorescence of the FH-probe without the complementary strand from the fluorescence signal in reactions containing the complementary target. The Tm was defined as the temperature at which 50% of subtracted fluorescence was reached.

2.4. PCR and signal to noise ratios

DNA qPCR was performed as described previously[15] using HotStarTaq Master Mix (Qiagen), 1uM of each primer and 0.1uM FH-probe. Experiments were done in duplicate using the real-time thermocycler IQ5 (BioRad, Basel) and as cycling profile: 95°C 15', 60x (95°C 5'', 55°C 5'', 60°C 40''). The following primers were used; cr1 (TCTCTGGCTAACTAGGGAACCCACTGCTT)[16] and cr2 (TGACTAAAAGGGTCTGAGGGATCTCTAGTTACCAG)[16] for the *early* region, ts5'gag (CAAGCAGCCATGCAAATGTTAAAAGA)[15] and skcc (TACTAGTAGTTCCTGCTATGTCACTTCC)[17] for the *gag* region, mf209 (AAAGCGTCTAGCCATGGCGTTAGTA) and mf302 (CAAATTTCTACTAATGCTTTTATTTTTTC) for the *pol* region, mf1 (CTTAGGCATCTCCTATGGCAGGAA)[18] and mf238 (GCTATTATTGCTGCTACTACTAATGCTACTA) for the *tat* region, mf222 (GGCAGGGATATTCACCATTATCGTTTCAGA) and mf83 (GGATCTGTCTCTGTCTCTCTCTCCACC)[18] for the *sa7* region, and mf345 (AATCAGGGAAGTAGCCTTGTGT) and mf346 (GAGGTGGGTTTTCCAGT) for the *nef* region . HXB2 was chosen as a standard target for all the experiments and pHXB2[19] was linearized by digestion with Xho1. Concentration of the plasmid was quantified by spectrophotometry. Based on the length of the linearized plasmid (13000bp), the molecular weight ($8.5x10^6$ g/mol) and copy number were calculated. A standard dilution series with 10fold dilution steps was prepared.

SNRs of individual FH-probes were determined by amplification of constant copy numbers ($3x10^6$ copies HXB2 DNA) dividing mean fluorescence of the last 5 cycles of the amplification when fluorescence has reached a plateau, by mean fluorescence of the initial 5 PCR cycles. Means of two independent experiments using duplicate measurements were calculated.

2.5. Calculations and Statistics

Statistical analyses were performed either using GraphPad Prism5.0 software (GraphPad Software, San Diego, CA) or Stata (Version 10.0; StataCorp).

Univariable and multivariable linear regression to determine predictors for SNRs of FH-probes were performed using Stata. $P<0.05$ was considered statistically significant. Factors with strong colinearity ($p<0.01$) were not included in the multivariable analysis.

The following parameters were tested in univariable analyses additionally to the ones shown in Table 2, but without revealing statistically significant association with SNRs: G/C-basepairs in 1st and 2nd position, purine in 1st and 3rd position, concentration of the FH-probe, Tm of the primer binding the opposite strand as the FH-probe and distance of FH-probe to the primer binding the same strand as the FH-probe.

3. Results

3.1. Phylogenetic complexity of conserved regions of the HIV-1 genome

Amplica in 6 regions commonly used for HIV-1 qPCR due to their conservation or usefulness for monitoring important splice variants were assessed in this study: An amplicon mapping the primary viral transcripts in the R/U5 region (*early*), the coding region for *p24-gag* (*gag*), the start of the reverse transcriptase gene (*pol*), the first coding region of the *tat*-gene (*tat*), *env-gp41* flanking the major splice acceptor 7 (*sa7*) and the poly-purine tract within the *nef* gene (*nef*) (Table 1, Fig. 1a).

Phylogenetic conservation within the six chosen amplica was analyzed with emphasis on match frequency, defined as 100% identity of FH-probe with its target, by employing a novel sequence scanning program. In this algorithm, a sequence defined as standard is screened in a sliding window approach for match frequencies of all possible FH-probes by comparison to a sequence database. In the present study, HXB2, the commonly used first molecular isolate of HIV-1[19], was used as the reference sequence and the database was a collection of 1242 near full-length HIV-1 genome sequences representing phylogenetically distinct viral isolates spanning the full scope of the HIV-1 pandemic, which was extracted from the Los Alamos HIV Sequence Database (hiv.lanl.gov). Figure 1b shows an analysis of 10-, 20- and 40-base oligomers plotted against average single base conservations. Theoretically, match frequencies (M) are proportional to the product of single base frequencies; this can be approximated by

equation 1) $M = (C_{GM})^{L}$,

which potentiates the geometric mean of single base conservation (C_{GM}) with oligonucleotide chain length (L).

A comparison of calculated versus observed match frequencies revealed close correlation (Pearson r=0.96) but significantly higher values for observed match frequencies (Wilcoxon signed rank test, $p<0.0001$). The most likely explanation for this discrepancy is that single base conservations were not independent from each other but genetically linked due to constraints of codon usage and RNA structure. Thus, to adjust the model for linkage, a correction factor (λ) was introduced:

Equation 2) $M = (C_{GM})^{L \times \lambda}$

Fitting the dataset to equation 2 resulted in λ=0.74 (95% confidence interval: 0.73-0.75). The model showed significant correlation (Pearson r=0.97) between observed

Table 1. Properties of FH-probes

			position[C]			match (%)[D]				Tm[F]	
ID[A]	FH-probe sequence[B]	amplica	5'	3'	length	all	nonB	B	SNR[E]	m	p
mf320	AAAGCTTGCCTTGAGTGCTTCA	early	530	551	22	36	29	52	4.4	64.5	67.7
mf321	TTCAAGTAGTGTGTGCCCGTCT	early	548	569	22	36	29	49	3.8	64.5	68.9
mf322	TGCCCGTCTGTTGTGTGACT	early	561	580	20	64	50	92	6.6	63.5	69.1
mf323	CCCGTCTGTTGTGTGACT	early	563	580	18	64	50	92	5.9	57.5	64.3
mf74	AGCACTCAAGGCAAGCTTTATTGAGGC [24]	early	548	522	27	87	82	97	3.9	64	72.1
ca20	TGtGtGCcCgT	early	557	567	11	83	77	96	12.9	59	68
ca21	TGtGtGccC	early	557	565	9	95	92	100	9.9	57	64
ca22	TGTGtGCCCGT	early	557	567	11	83	77	96	10.5	52	56
ca23	AGcTtGcCtTGaG	early	532	544	13	96	95	97	10.7	62	67
ca24	AGCttGCCttGaG	early	532	544	13	96	95	97	27.3	62	70
ca25	AGCtTGCCtTGAG	early	532	544	13	96	95	97	8.2	53	54
mf316	TAGAGTGCATCCAGTGCATGCA	gag	1428	1449	22	0.8	0	4.8	5.8	65	68.7
mf317	ATGCAGGGCCTATTGCACCA	gag	1445	1464	20	22	16	51	6.5	63	68.8
mf318	AGGCCAGATGAGAGAACCAAGG	gag	1464	1485	22	24	15	65	4	69	68.9
mf319	TGCAGCTTCCTCATTGATGGT	gag	1419	1399	21	43	36	78	4.8	63	66.8
Boe3	TCTATCCCATTCTGCAGCTTCCTCATT [25]	gag	1431	1405	27	15	3.3	70	1.9	70.5	70.6
Boe3.2	TCTATCCCATTCTGCAGCTTCCTCATTGATGG	gag	1431	1400	32	13	2.8	66	1.7	73	74.3
Boe3.3	TCCCATTCTGCAGCTTCCTCATTGATGG	gag	1427	1400	28	38	32	69	2	56.3	73.5
ri17	CCAtCaAtGaGGA	gag	1400	1412	13	68	64	88	5	56	59
ri18	CTGcAgAaTgGGA	gag	1415	1427	13	80	78	86	1.3	62	66
ri19	CATgCaGgGcCT	gag	1444	1455	12	59	57	70	4.2	63	72
ri20	AGGcCcTgc	gag	1455	1447	9	79	77	90	5.9	58.5	61
ri21	TCccattc	gag	1427	1420	8	87	87	89	1.2	nn	53
mf304	ATGGCCCAAAAGTTAAACAATGGCCA	pol	2599	2624	26	22	13	67	4.9	66	71
mf305	ATTCCTGGCTTTAATTTTACTGGTACAGT	pol	2596	2568	29	47	41	76	2.2	66.5	68.2
mf309	TTAAACAATGGCCATTGACAGAA	pol	2611	2633	23	63	60	75	2.3	57	64.2
mf348	AAGCCAGGAATGGATGGCC	pol	2586	2604	19	63	60	81	8.5	65.5	67
ri14	ACTgTaCcAgTA	pol	2568	2579	12	83	81	91	1.6	54	54
ri15	CAGgAaTgGaTGG	pol	2590	2602	13	87	86	95	9.5	59	62
ri16	CTGtCaAtGgCCA	pol	2631	2619	13	82	81	90	9.3	63.5	64
ri23	ATGgatgg	pol	2595	2602	8	97	96	98	1.5	53	54
mf82	TCTTCGTCGCTGTCTCCGCTTCTT [17]	tat	6001	5978	24	17	7.6	64	2.5	61.5	72.7
mf324	TCCCAACCCCGAGGGGA	sa7	8386	8402	17	2.7	0.8	12	2.8	64.5	69.8
mf325	CCCGAAGGAATAGAAGAAGA	sa7	8412	8431	20	5.4	5.9	2.9	3.9	58.5	61.2
mf326	CTCGGGGTTGGGAGGTG	sa7	8398	8382	17	0.6	0	3.4	2.9	63.5	66.6
mf327	CTATTCCTTCGGGCCTGTC	sa7	8424	8406	19	5.5	6	2.9	4.4	63.5	64.4
mf328	ACCCCGAGGGGACCC	sa7	8391	8405	15	2.7	1.1	11	4	62	67.2
mf2tq	TTCCTTCGGGCCTGTCGGGTCCC [17]	sa7	8421	8399	23	18	9.2	61	6.5	74.5	77.4
mf226	AGGGGACCCGACAGGCCC [15]	sa7	8397	8414	18	23	13	73	18.1	72	73.4
ri10	TGTcGgGtCcCCTC	sa7	8409	8396	14	32	24	74	3.8	73	75
ri11	CCTgTcgGGtCCC	sa7	8411	8399	13	38	29	79	3.2	59	73
ri12	CCCgACaGgCcCG	sa7	8403	8415	13	36	28	77	3.9	64	77
ri13	CGAcAgGcCCgAA	sa7	8405	8417	13	31	24	69	4.3	58	76
ri22	CGAcAggC	sa7	8405	8412	8	75	73	87	1.4	53	55
mf342	TTTTTAAAAGAAAAGGGGGGAC	nef	9064	9085	22	86	88	77	1.1	66.5	62.3
mf343	AAGAAAAGGGGGGACTGGA	nef	9071	9089	19	87	89	77	1.3	63	65.5
mf344	ACAGATCAAGGATATCTTGTCT	nef	9133	9112	22	8	2.4	36	1.7	54	61.4
mf341	CTTTTTAAAAGAAAAGGGG	nef	9063	9081	19	83	86	70	1	42	54.5
ca18	CAAGGCAGCTGTAGATCTTAGCCA	nef	9039	9062	24	3.4	0.5	18	3.8	63	68.8
ca19	AGGGCTAATTCACTCCCAACGA	nef	9090	9111	22	0.2	0.1	0.5	9.3	63	68.5
ca26	AGGGGgGaCtGgA	nef	9077	9089	13	95	95	95	4.2	58	74

[A] ID denominates identification of FH-probes according to internal nomenclature in the author's laboratory.
[B] Sequence of the FH-probe (5' to 3'): capital letters indicate DNA bases and lower case letters indicate LNA bases.
[C] Numbering based on the HXB2 genome according to the software provided by the Los Alamos HIV Sequence Database (SEQUENCE LOCATOR, www.hiv.lanl.gov).
[D] Conservation of all subtypes (all), nonB- (nonB) or B-subtypes (B) was calculated by comparison to sequences that were obtained from curated alignments in the Los Alamos HIV Sequence Database (download 2007, www.hiv.lanl.gov).
[E] Signal to noise ratio
[F] Measured (m) and predicted (p) melting temperature (Tm) (°C)

and predicted match frequency and no significant difference between them (Wilcoxon signed rank test, p=0.32). This analysis demonstrates that the probability that an oligonucleotide will find its match in a viral population decreases exponentially with FH-probe length (Fig. 1c). The influence of the oligonucleotide length may be relieved if genetic linkage results in coinheritance of clusters of bases, in which case, $\lambda<1$, oligonucleotide length is virtually shortened.

In contrast to these constraints related to assay sensitivity, specificity of FH-probe binding (S), defined as the probability that a certain sequence would not occur in a random amplicon, can be approximated by

equation 3) $S = (1-(1/4)^{L})^{(A-L+1)}$,

where L is the length of the FH-probe and A the length of a random, erroneously amplified sequence. This equation can be justified as follows: The probability, that the FH-probe matches a given stretch of length L (i.e. the length of the FH-probe) is $(1/4)^{L}$ (assuming equilibrium of all 4 bases). The entire amplicon can be viewed as consisting of A-L+1 stretches of length L. This approximation further assumes that the event of matching a given stretch is independent from matching the other stretches. Under this assumption the probability of obtaining no match at all is $(1-(1/4)^{L})^{(A-L+1)}$. Thus, it was estimated that the specificity of FH-probe binding exceeds 99% with a FH-probe length of $\geq$9bases and A ranging from 40-1000 bases (Fig. 1c).

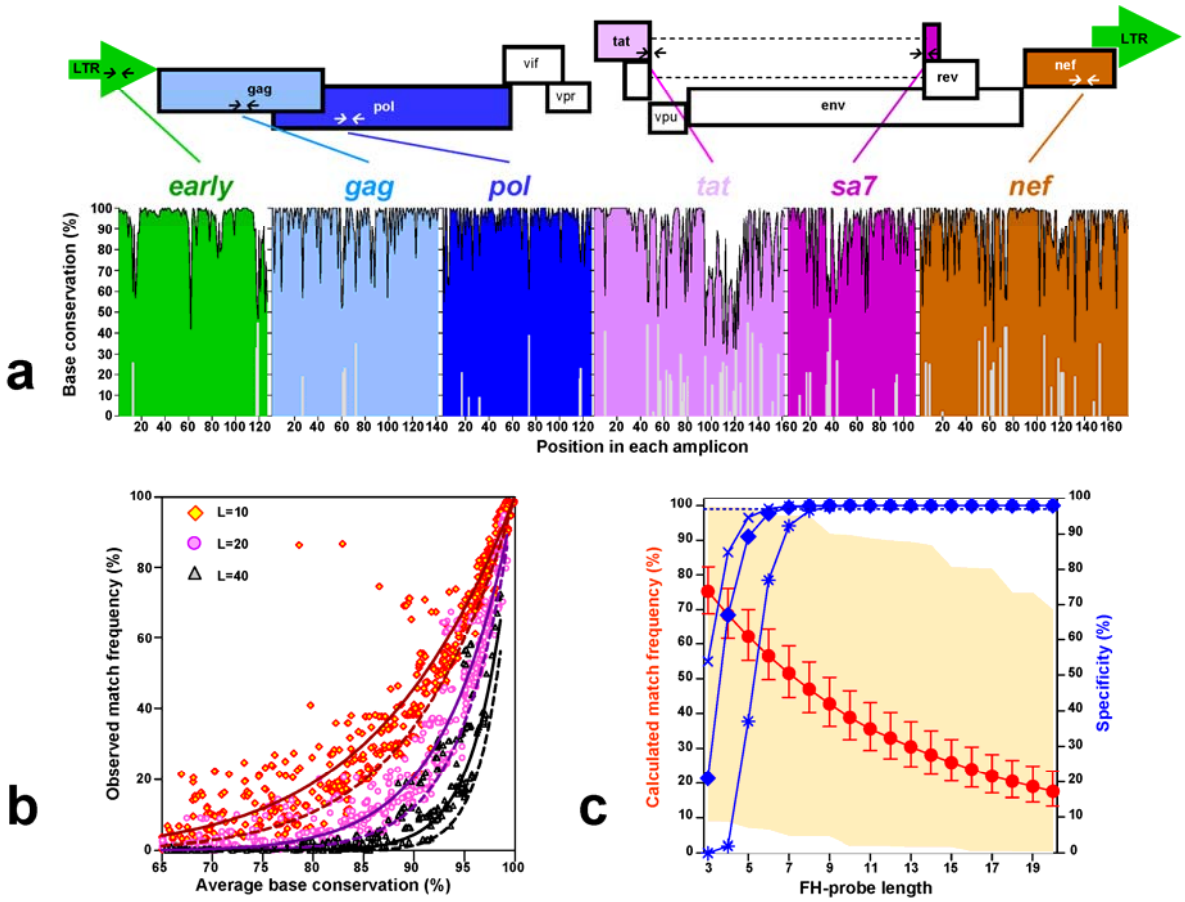

Figure 1. Phylogenetic complexity of conserved regions of the HIV-1 genome.(a) FH-probes mapping to 6 regions of the HIV-1 genome were designed utilizing the HXB2 sequence as a template. Upper panel: Small arrows mark the location of amplica in the viral genome. Lower panel: Conservation of the chosen amplica among all circulating HIV-1 clades (subtypes A-O, sequences obtained from curated alignments in the Los Alamos HIV Sequence Database, download 2007, www.hiv.lanl.gov). The colored area shows the level of single base conservation as compared to a global consensus sequence, and grey bars signify the conservation of the actual HXB2-sequence when this deviates from consensus. (b) Dependence of match frequencies of oligonucleotides on base conservation and oligonucleotide-length (L). HIV-1 sequence alignments were scanned for the presence of all possible 10- (red, yellow), 20- (magenta), and 40-mers (black) in the 6 chosen amplica. Match frequencies (M) were plotted against average (GM, geometric mean) of single base conservation (C). Broken lines show prediction of M by equation 1) $M = (C_{GM})^{L}$. Steady lines depict predictions by equation 2) $M = (C_{GM})^{Lx\,\lambda}$ including a correction factor (λ) to account for linkage. Brown lines show predictions for 10-mers, magenta lines for 20-mers and black lines for 40-mers. (c) Calculated match frequencies (M according to equation 2) (left y-axis; red circles, error bars: 95% confidence intervals, shaded area: range). Specificity (S, right y-axis) of FH-probe binding is modelled as a function of FH-probe length (L) and length of an erroneously produced amplicon (A): Equation 3) $S = (1-(1/4)^{L})^{(A-L+1)}$ (A=40 bases: blue crosses, A=100 bases: blue diamonds, A=1000 bases: blue asterisks)

3.2. Signal to noise ratios of FH-probes as indicators of DNA FH-probe performance

In previous studies signal to noise ratios (SNRs) of fluorescent probes were used to gauge efficacy of qPCR assays[20,21]. To confirm this concept, SNRs of individual FH-probes were calculated by dividing fluorescence at the end of the amplification by fluorescence during the initial cycles of amplification. To avoid interference of qPCR by possible primer dimer amplification and the resulting reduction of fluorescence in reactions with low copy numbers[22], high amounts of template ($3x10^6$ copies HXB2 DNA) were used. As shown in Figure 2a and Table 1, performance in a panel of 30 DNA-based FH-probes ranged from ineffective (mf342, SNR=1.1) to excellent (mf226, SNR=18.7). Cycle threshold (C_t) -values and their standard deviations, respectively, showed significant inverse correlation with SNRs (Pearson r=-0.52; p=0.004 and r=-0.59; p=0.0007 respectively), confirming association of SNRs with sensitivity and reproducibility of qPCR.

3.3. Predictors of SNRs of DNA FH-probes

To assess the factors influencing biochemical performance of the DNA FH-probes, SNRs and their different potential predictors were examined in univariable and multivariable linear regression models (Table 2). Melting temperature (Tm) showed weak but significant correlation to SNRs, either when measured or predicted by nearest neighbour analysis (p=0.033; measured and p=0.038; predicted). High G/C content of the DNA FH-probe was associated (p=0.002) with high SNRs. Furthermore, the presence of G/C rich secondary structures within the FH-probe was, in contradiction to current rules, a highly significant positive predictor of SNRs (p<0.001). FH-probes with intramolecular stem-loop structure containing one or more G/C base-pairs performed better than FH-probes with no or only a weak secondary structure (Fig. 2b, p=0.003). A trend for association of short FH-probe length (p=0.102) with high SNRs was observed, while Tm-difference of FH-probe and the primer binding the same strand (Delta-Tm) lacked correlation (p=0.742).

Upon encountering FH-probe bound to its template, taq polymerase can either remove the FH-probe by exonucleolytic cleavage or by strand displacement. We hypothesized that stability of the probe-template complex at its 5' end may favour exonucleolytic cleavage and generation of free, unquenched fluorophore. Thus,

thermal stability and base composition of the 5' ends of the FH-probes were analyzed in detail.

SNRs were positively associated with the predicted Tm of the ten 5' proximal bases (Pearson r=0.65, p=0.0001) (Fig. 2c). In agreement, FH-probes with a G or C at position 3 showed significantly elevated SNRs as compared to those with A or T (p=0.001, data not shown). At the second position, presence of G was favourable, followed by presence of A. Hence, SNRs of DNA FH-probes with a purine at position two were significantly higher than SNRs with a pyrimidine (Fig. 2d, p=0.01). No significant association of the bases at the first position with SNRs was observed.

Multivariable analysis, including statistically significant factors (p≤0.05 in univariable analysis), except for those with strong colinearity (p<0.01), resulted in a model in which 5'-proximal Tm (p=0.006), a purine at position 2 (p=0.035) and the numbers of intramolecular G/C-basepairs (p=0.004), explained 67% of the variability in SNRs (Table 2).

Thus, we suggest the following rules to design functional DNA FH-probes: 5'-proximal Tm ≥ 30°C, a purine at position 2 and ≥ 2 intramolecular G/C-basepairs.

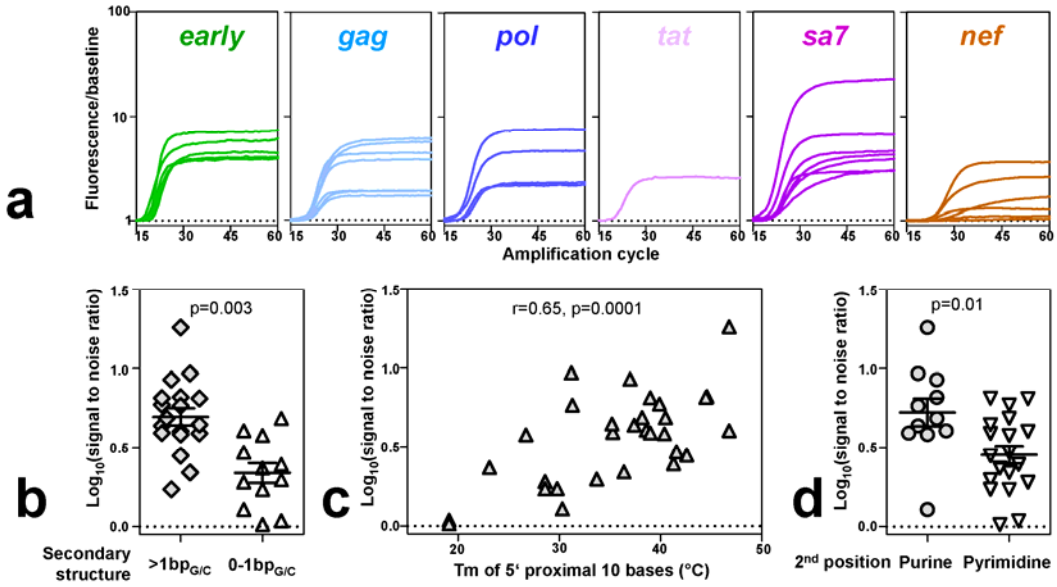

Figure 2. Performance of DNA FH-probes and analysis of predictors of SNRs. (a) Amplification curves of DNA FH-probes in 6 amplica (n=1-7, see Table 1). Fluorescence at each cycle was normalized to the baseline fluorescence of the first 5 cycles. Dotted lines indicate baseline fluorescence. (b) Comparison of SNRs of DNA FH-probes containing <1 (diamonds) or 0-1 G/C (triangles) base-pairs in a stem-loop. (c) Pearson correlation analysis of the Tm of the 5' proximal 10 bases of the DNA FH-probes with SNRs. (d) Comparison of DNA FH-probes containing a purine (circles) or a pyrimidine (triangles) at the 2nd position. P-values in (b) and (d) were calculated by Mann Whitney testing.

3.4. Performance of DNA FH-probes in qPCR

Four FH-probes with SNR>2 were chosen for further analysis according to their phylogenetic match frequencies: mf74 within *early* (87% match to viral isolates from all clades, 97.5% clade B), mf319 in *gag* (42% overall, 77.5% clade B), mf348 in *pol* (63% overall, 81% clade B), and mf226 in *sa7* (22.8% overall, 72.6% clade B). FH-probes within *nef* and *tat* were not included due to low match frequencies within the set of utilizable FH-probes (<20%). HIV-1 DNA copies were measured by qPCR in a range from 3 million to 0.3 copies per PCR and the 50% endpoint of PCR-positive dilutions was determined as an indicator of assay-sensitivity. All qPCRs using *gag*, *pol* and *sa7* FH-probes reached single copy sensitivity (Fig. 3b-d), while amplification of the *early* amplicon showed slightly reduced sensitivity presumably due to formation of primer dimers in later cycles (Fig. 3a), typically resulting in progressive diminution of signal amplitudes[15].

Thus, quantification of HIV-1 DNA was achieved over 6 orders of magnitude approaching single copy sensitivity for 3 FH-probes. Although they matched B-subtypes with reasonable frequency, match frequencies for nonB-subtypes were insufficient to cover the full scope of the global pandemic.

3.5. Improved short FH-probes using locked nucleic acid analogues

To design FH-probes with higher clade coverage, shorter chimeric DNA locked nucleic acid (DLNA) FH-probes were used. In locked nucleic acids base pairing is stabilized by a 2-O,4-C-methylene bridge on the ribose moiety[23]. As exemplified in Figure 4a, a 13-mer DLNA FH-probe showed higher thermal stability when bound to its complementary strand than an overlapping 23mer DNA FH-probe.

To identify candidate short FH-probes with high conservation, the sequence scanning algorithm was employed. An example is shown for the *pol* amplicon (Fig. 4b), which was scanned for all possible 8-, 10-, 12- and 14-mers. This procedure identified separate peaks of highly conserved sequence stretches from which FH-probes were chosen. Based on this analysis, 21 DLNA FH-probes (6 in *early*, 5 in *gag*, 4 in *pol*, 5 in *sa7*, 1 in *nef,* Table 1) ranging from 8-14 bases were synthesized.

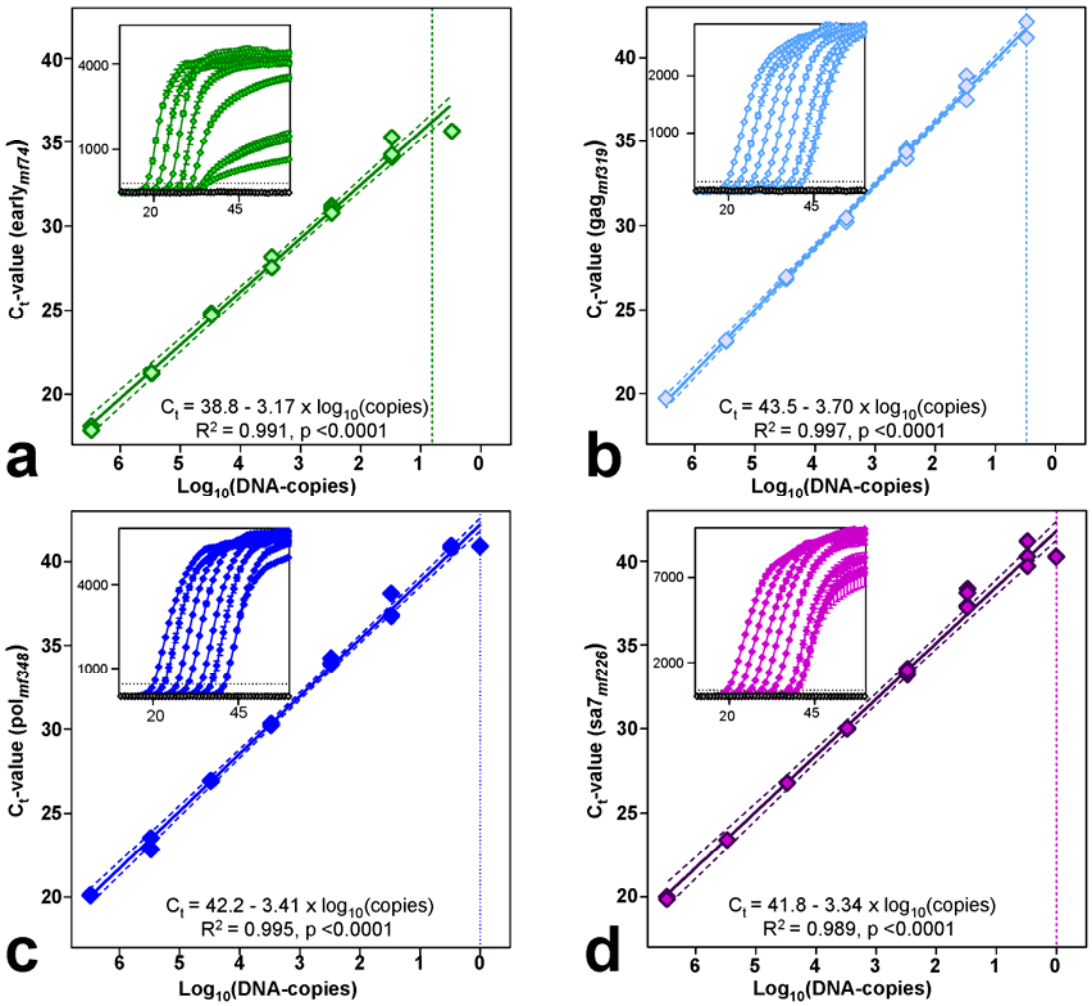

Figure 3. Sensitivity of qPCR using selected DNA FH-probes. DNA FH-probes with SNRs > 2 and the highest phylogenetic match frequencies of HIV-1 in each amplicon were used for amplification of HXB2 DNA in a range from 3 million to 0.3 copies per PCR. Broken vertical lines show the 50% endpoint of PCR-positive dilutions and thus signify sensitivity of the assays. qPCR for *gag* (b), *pol* (c) and *sa7* (d) reached single copy sensitivity, while amplification of the *early* (a) amplicon showed reduced sensitivity presumably due to primer dimer formation. Experiments were performed in duplicate ($3x10^6$-$3x10^3$ copies) or quadruplicate reactions (<$3x10^3$ copies). Results of linear regression of $\log_{10}$ of DNA copy numbers versus C_t-values are depicted. Inserts show amplification plots of PCR-cycle (x-axis) versus baseline subtracted fluorescence (y-axis), with dotted lines displaying the fluorescence threshold and grey symbols showing negative controls.

3.6. Predictors of SNRs of DLNA FH-probes

SNRs of DLNA FH-probes were measured as described for DNA FH-probes. C_t-values and their standard deviations, respectively, were inversely correlated with SNRs (Pearson r=-0.55; p=0.01 and Pearson r=-0.73; p=0.0003).

SNRs of DLNAs were tested for association with Tm, FH-probe length, G or C at position 3 and purine at position 2. Intramolecular G/C-basepairs were not assessed because DLNA FH-probes generally contained no secondary structures. Similarly, the Tm of the ten 5' proximal residues was not assessed as it was virtually identical to overall Tm. Finally, to account for the fact that not only G/C-residues promote strong basepairing but also LNA-residues, the presence of weakly pairing bases (plain A/T content, i.e. no G/C no LNA) was assessed rather than presence of strongly pairing bases (G/C, LNA).

The sole statistically significant predictor identified in univariable analysis was purine at position 2 (p=0.009). For identification of additional potential predictors, inclusion in the multivariable model was extended to parameters with a univariable p-value ≤0.1. This resulted in a model explaining 56% of the data variability (Table 2), in which purine at position 2 (p=0.007) and FH-probe length (p=0.023) positively predicted SNRs, while plain A/T-content showed a negative association (p=0.040).

Thus, we propose the following rules to design functional DLNA FH-probes: A purine at position 2, length ≥ 11 nucleotides and a plain A/T-content ≤ 35%.

Table 2. Predictors of SNRs of DNA and DLNA FH-probes

		Univariate analyses			Multivariate analyses			
		Coefficient (95% CI)		p	Coefficient (95% CI)		p	r^2
DNA	Tm measured (°C)[A]	0.018	(0.001; 0.03)	0.038	0.005	(-0.01; 0.02)	0.356	0.68
(n=30)	G/C content (%)[B]	0.013	(0.005; 0.02)	0.002				
	FH-probe length	-0.022	(-0.05; 0.005)	0.102				
	Predicted Tm 5' end (°C)[C]	0.025	(0.01; 0.04)	<0.0001	0.015	(0.005; 0.03)	0.006	
	G/C-basepair in 3rd position	0.331	(0.14; 0.52)	0.001				
	Purine in 2nd position	0.27	(0.07; 0.47)	0.01	0.155	(0.01; 0.30)	0.035	
	Delta-Tm (°C)[D]	0.003	(-0.01; 0.02)	0.742				
	Intramolecular G/C-basepairs	0.12	(0.06; 0.17)	<0.0001	0.074	(0.03; 0.12)	0.004	
DLNA	Tm measured (°C)[A]	0.011	(-0.02; 0.05)	0.53				0.56
(n=21)	Plain AT-content (%)[E]	-0.014	(-0.03; 0.002)	0.091	-0.014	(-0.03;-0.001)	0.04	
	FH-probe length	0.068	(-0.02; 0.15)	0.107	0.077	(0.01; 0.14)	0.023	
	G/C-basepair in 3rd position	0.044	(-0.31; 0.40)	0.797				
	Purine in 2nd position	0.423	(0.12; 0.73)	0.009	0.377	(0.12; 0.64)	0.007	
	Delta-Tm (°C)[D]	-0.009	(-0.03; 0.02)	0.453				

[A] Measured melting temperature. [B] Content (%) of G and C bases. [C] Predicted melting temperature of the 10 proximal bases at the 5' end. [D] Difference between the Tm of the FH-probe and the Tm of the primer binding the same strand as the FH-probe. [E] Content (%) of A and T bases, not including LNA bases

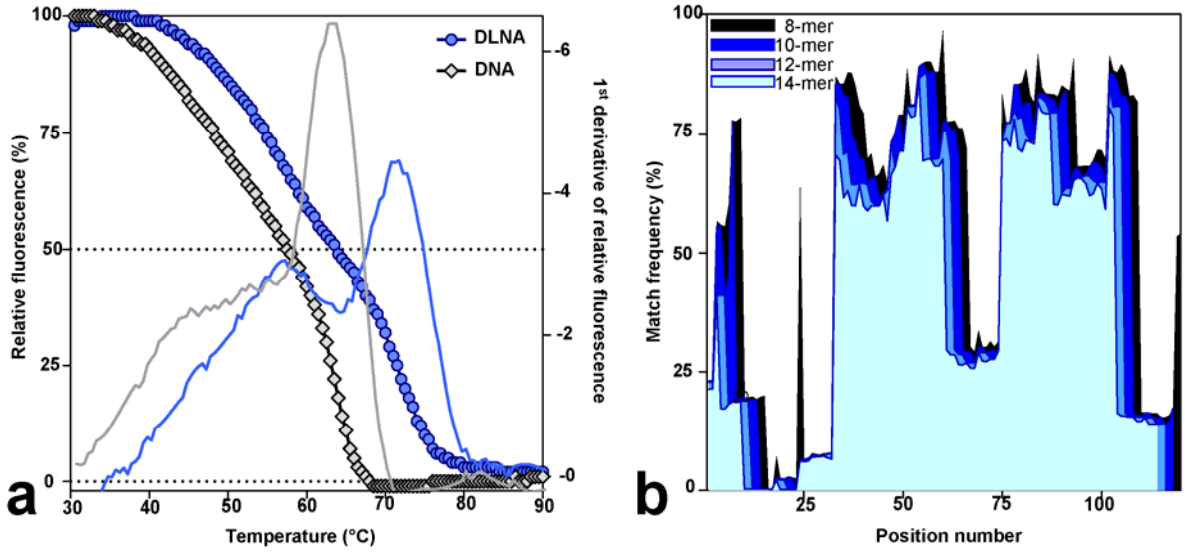

Figure 4. Melting profiles and match frequencies of DLNA FH-probes. (a) Melting profiles of two FH-probes within *pol*. The 13-mer DLNA FH-probe ri16 is depicted in blue and the overlapping 23-mer DNA FH-probe mf309 in grey. Symbols connected by lines show relative fluorescence and plain lines show their first derivatives. (b) Matching of HXB2 derived 8-, 10-, 12- and 14-mers to the aligned HIV-1 sequences of the Los Alamos HIV Sequence Database (download 2007, www.hiv.lanl.gov) starting at the indicated position within the *pol* amplicon (position 2500 to 2630, HIV-1 HXB2, GenBank accession number K03455).

3.7. Performance of DLNA FH-probes in qPCR

Based on their SNRs and phylogenetic conservation, DLNA FH-probes representing *gag (*ri20, 79% match to viral isolates from all clades, 90% in subtype B), *sa7* (ri12, overall 36%, subtype B 77%), *nef* (ca26, 95% overall and for clade B) and 2 FH-probes for *pol* (ri15, 87% overall, 94% clade B and ri16, 82% overall, 90% clade B) were further tested for their usefulness in qPCR (Fig. 5). FH-probes for the *early* amplicon were not included in this analysis because of the slightly reduced sensitivity of the qPCR, as shown above for DNA FH-probes, and because the dataset for this amplicon compassed a smaller number of viral variants than the other amplica (n=233 versus n=1242) with overrepresentation of subtype B (34% versus 17%). All PCR assays ranged over 6 orders of magnitude and reached single copy sensitivity.

In a previous study match frequencies of an RT-qPCR assay were boosted by adding two FH-probes into a single reaction[18]. The rationale was that fluorescence is expected to show up when at least one of two FH-probes A and B fits its target. This combined match frequency (M_{AvB}) can be estimated as

equation 4) $M_{AvB} = M_A + M_B - (M_A \times M_B)$.

This concept was employed with 2 FH-probes in *pol* (ri15 and ri16), which resulted in qPCR with single copy sensitivity (Fig. 5d). Assays performed with single FH-probes gave indistinguishable results (data not shown). Combinatorial analysis as described by equation 4, predicted that 97.8% (observed 96.7%) of viral isolates among all clades and 99.4% (observed 99.0%) within clade B, would match at least one of these two FH-probes. Thus, combination of two *pol* FH-probes resulted in near-universal clade coverage.

The *nef* DLNA FH-probe (ca26) showed the highest match frequency (95%) and good performance in DNA assays. However, when it was employed for RNA in one-step RT-qPCR under less stringent conditions (annealing at 50°C for 30 minutes), false positive reactions were observed at low frequencies presumably due to the stretch of 6 guanosines in a row. The other 4 DLNA FH-probes in this subset functioned reproducibly, with great specificity and without giving rise to false positive results using uninfected human PBMC DNA as a template, as well as in RT-qPCR (unpublished results).

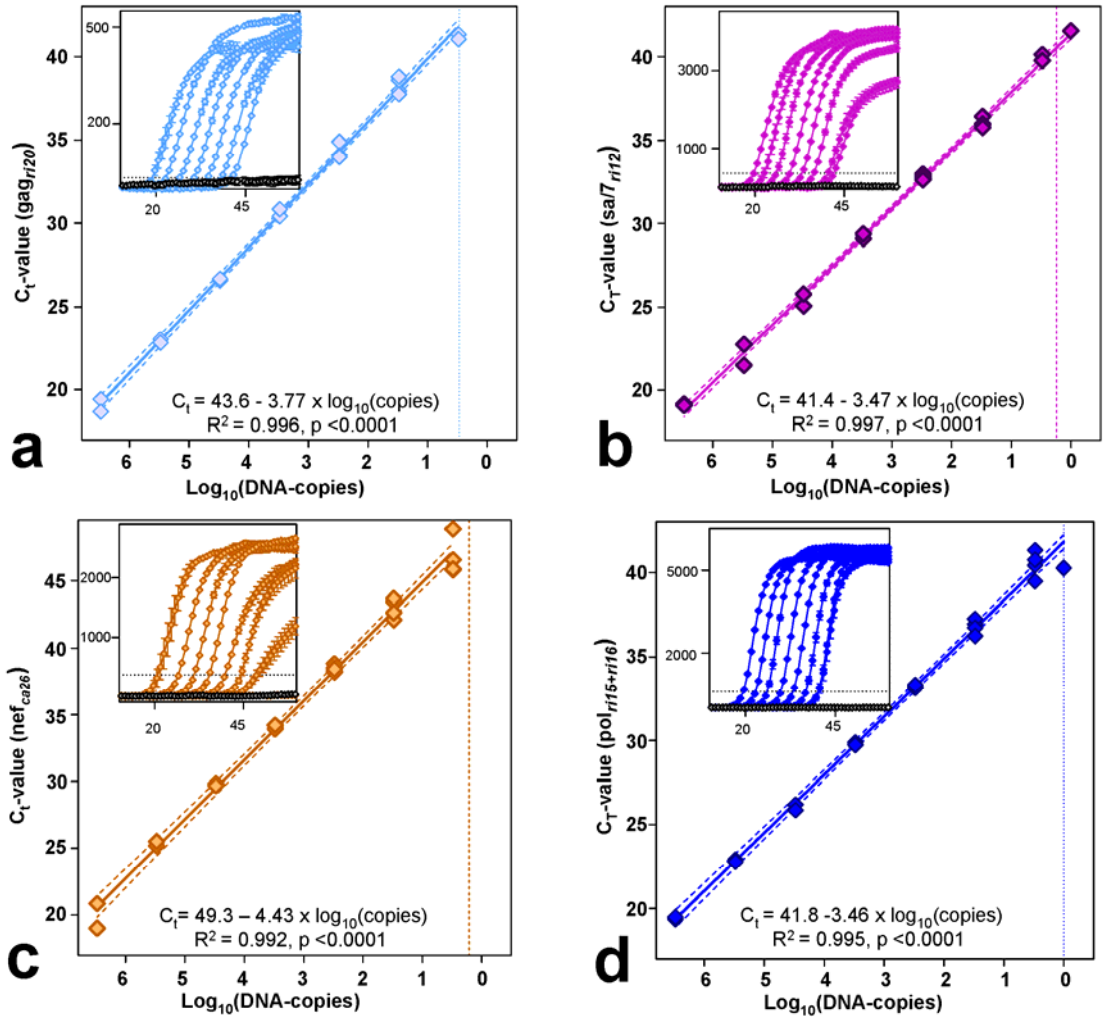

Figure 5. Sensitivity of qPCR using selected DLNA FH-probes. DLNA FH-probes with SNRs > 2 and the highest phylogenetic match frequencies of HIV-1 in each amplicon were used for amplification of HXB2 DNA in a range from 3 million to 0.3 copies per PCR. Broken vertical lines show 50% endpoints of PCR-positive dilutions and thus signify sensitivity of the assays. All qPCRs using *gag* (a), *sa7* (b), and *nef* (c) DLNA FH-probes reached single copy sensitivity. Two *pol* FH-probes that had both reached single copy sensitivity in qPCR on their own, were combined in one qPCR (d). This assay also reached single copy sensitivity. Experiments were performed in duplicate ($3x10^6$-$3x10^3$ copies) or quadruplicate reactions (<$3x10^3$ copies). Results of linear regression of $\log_{10}$ of DNA copy numbers versus C_t-values are depicted.
Inserts show amplification plots of PCR-cycle (x-axis) versus baseline subtracted fluorescence (y-axis), with dotted lines displaying the fluorescence threshold and grey symbols showing negative controls.

4. Discussion

In the present study, both the structural as well as the phylogenetic side of FH-probe design for HIV-1, a paradigm for a virus with high genetic diversity, were addressed. In consideration of the finding that match frequencies decrease exponentially with sequence length, usage of short optimally conserved FH-probes was aimed for. This has been facilitated by a novel computational tool to systematically identify stretches of highly conserved regions to which FH-probes may be targeted. This and the availability of locked nucleic acid analogues enabled us to design short FH-probes matching up to 95% of a representative collection HIV-1 isolates. Similarly, by combining two FH-probes within the same reaction, phylogenetic coverage could even be increased to 97%.

Regarding performance of FH-probes, analysis of the dataset in this study allowed to determine rules empirically for FH-probe design with the limitation that the collection of FH-probes was not strictly random and that certain rules were initially followed in design of FH-probes.

In agreement with current rules[6], a significant impact of stability of the probe-target complex both on DNA as well as on DLNA FH-probes was observed. However, more than the Tm, representing overall stability, the local stability near the site of nucleolytic cleavage of the FH-probe appeared to exert a marked influence on FH-probe performance. Hence, the Tm of the first 10 5'-proximal bases was a significant positive predictor of the SNRs. Moreover, stabilization of the probe-target duplex appeared to be mediated by forces beyond mere Watson Crick basepairing, as indicated by the observation that a purine at position 2 significantly predicted SNRs in the two independent datasets comprising DNA and DLNA FH-probes. This finding may reflect hydrophobic interaction of the FH-probe fluorophore with the extensive hydrophobic structure of the purine-rings.

Paradoxically, presence of intramolecular G/C base-pairs within DNA FH-probes was significantly associated with high SNRs despite their potential to compete with hybridization to the target strands. In analogy to molecular beacon probes which rely on stem-loops bringing fluorophore and quencher in juxtaposition[24], intramolecular G/C base-pairs in FH-probes may increase quenching in the absence of target. In agreement, significant correlation of intramolecular G/C base-pairs with baseline fluorescence (data not shown, Pearson $r=-0.53$, $p=0.003$) was observed.

Similarly to DNA FH-probes, the three identified predictors of DLNA FH-probe performance, purine 2, chain-length and low content of plain A/T base-pairs, all may influence duplex-stability. The fact that the Tm appeared not to predict FH-probe performance can be interpreted as a consequence of sampling, since DLNA FH-probes were chosen to exceed, whenever possible, a predicted Tm of 55°C in order to have a chance to perform well.

5. Conclusions

In summary, the following empirically tested positive predictors of biochemical FH-probe performance emerged from this study:

1. A purine at position 2 (DNA and DLNA).
2. An overall Tm ≥ 60°C and Tm of the first ten 5' proximal bases ≥ 30°C (DNA).
3. Presence of G/C rich (≥ 2 G/C-basepairs) secondary structures (DNA).
4. Sequence length (≥ 11) and low contents (≤ 35%) of plain A/T bases (DLNA).

Moreover, within the set of well performing FH-probes, application of the algorithm to scan sequence databases for FH-probes with optimal phylogenetic conservation, allowed to identify functional FH-probes in various regions of the HIV-1 genome approaching coverage of the global HIV-1 pandemic. It is conceivable that application of these phylogenetic and biochemical principles for FH-probe design may also be extended to other phylogenetically diverse biological systems.

Acknowledgements

This work was supported by a grant from the „Stiftung für Wissenschaftliche Forschung an der Universität Zürich“ to MF and by the Swiss National Science Foundation (grant no. 3100A0-112670 to MF and grant no. 324730-116035 to HG).

We gratefully acknowledge Gertie Heimbeck (Biorad), for her skilful comments on real-time PCR and maintenance of real-time thermocyclers and Alexandra Trkola for her help in the initiation of the study and for helpful discussions.

References

1. Christopherson, C., Sninsky, J., and Kwok, S., The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. Nucleic Acids Res 25 (3), 654 (1997).
2. Damond, F. et al., Human immunodeficiency virus type 1 (HIV-1) plasma load discrepancies between the Roche COBAS AMPLICOR HIV-1 MONITOR Version 1.5 and the Roche COBAS AmpliPrep/COBAS TaqMan HIV-1 assays. J Clin Microbiol 45 (10), 3436 (2007).
3. Ranade, K. et al., High-throughput genotyping with single nucleotide polymorphisms. Genome Res 11 (7), 1262 (2001).
4. Livak, K. J., Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal 14 (5-6), 143 (1999).
5. Gut, M., Leutenegger, C. M., Huder, J. B., Pedersen, N. C., and Lutz, H., One-tube fluorogenic reverse transcription-polymerase chain reaction for the quantitation of feline coronaviruses. J Virol Methods 77 (1), 37 (1999).
6. Mackay, I. M., Arden, K. E., and Nitsche, A., Real-time PCR in virology. Nucleic Acids Res 30 (6), 1292 (2002).
7. Bruijnesteijn Van Coppenraet, E. S. et al., Real-time PCR assay using fine-needle aspirates and tissue biopsy specimens for rapid diagnosis of mycobacterial lymphadenitis in children. J Clin Microbiol 42 (6), 2644 (2004).
8. Malnati, M. S. et al., A universal real-time PCR assay for the quantification of group-M HIV-1 proviral load. Nat Protoc 3 (7), 1240 (2008).
9. Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W., and Deetz, K., Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl 4 (6), 357 (1995).
10. Letertre, C., Perelle, S., Dilasser, F., Arar, K., and Fach, P., Evaluation of the performance of LNA and MGB probes in 5'-nuclease PCR assays. Mol Cell Probes 17 (6), 307 (2003).
11. Proudnikov, D. et al., Optimizing primer--probe design for fluorescent PCR. J Neurosci Methods 123 (1), 31 (2003).
12. Lunge, V. R., Miller, B. J., Livak, K. J., and Batt, C. A., Factors affecting the performance of 5' nuclease PCR assays for Listeria monocytogenes detection. J Microbiol Methods 51 (3), 361 (2002).
13. Hall, T. A., Bioedit: A user-friendly biological sequence alignment editor and analysis program for Windows 95=98 NT. Nucleic Acids Symp Ser 41, 95 (1999).
14. Ihaka, R. and Gentleman, R., R: a language for data analysis and graphics. Journal of computational and graphical statistics, 299 (1996).
15. Kaiser, P. et al., Productive Human Immunodeficiency Virus 1 Infection in peripheral Blood Predominantely Takes Place in CD4/CD8 double negative T Lymphocytes. J Virol 81, 9693 (2007).
16. Lewin, S. R. et al., Use of Real-Time PCR and molecular beacons to detect virus replication in human immunodeficiency virus type 1-infected individuals on prolonged effective antiretroviral therapy. J.Virol. 73, 6099 (1999).
17. Christopherson, C. et al., PCR-Based assay to quantify human immunodeficiency virus type 1 DNA in peripheral blood mononuclear cells. J.Clin.Microbiol. 38 (2), 630 (2000).

18. Fischer, M. et al., Cellular viral rebound after cessation of potent antiretroviral therapy predicted by levels of multiply spliced HIV-1 RNA encoding nef. J Infect Dis 190 (11), 1979 (2004).
19. Ratner, L. et al., Complete Nucleotide-Sequences of Functional Clones of the Aids Virus. Aids Research and Human Retroviruses 3 (1), 57 (1987).
20. Yao, Y., Nellaker, C., and Karlsson, H., Evaluation of minor groove binding probe and Taqman probe PCR assays: Influence of mismatches and template complexity on quantification. Mol Cell Probes 20 (5), 311 (2006).
21. Kutyavin, I. V. et al., 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids Res 28 (2), 655 (2000).
22. Kaiser, P. et al., Equal amounts of intracellular and virion-enclosed hepatitis C virus RNA are associated with peripheral-blood mononuclear cells in vivo. J Infect Dis 194 (12), 1713 (2006).
23. Jensen, G. A., Singh, S. K., Kumar, R., Wengel, J., and Jacobsen, J. P., A comparison of the solution structures of an LNA : DNA duplex and the unmodified DNA : DNA duplex. Journal of the Chemical Society-Perkin Transactions 2 (7), 1224 (2001).
24. Tyagi, S. and Kramer, F. R., Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 14 (3), 303 (1996).

CHAPTER 4

Association between specific HIV-1 Env traits and virologic control *in vivo*

Published as B. Joos *et al.* / Infection, Geneteics and Evolution 10 (2010) 365-372

Own contribution

After primary HIV infection each individual reaches a steady state level of plasma viral load which is referred to as viral setpoint. These setpoints between individuals can vary more than 1000 fold and are predictive for disease progression to AIDS. In chronically infected patients, who underwent structured treatment interruptions (STI), low viral diversity before initiation of ART, was associated with spontaneous control of viremia after treatment stop. In addition, viral capacity is a driving factor in determining the magnitude of the viral setpoint. Taken together, viral properties influence disease progression and are at least in part responsible for the high variability in the viral setpoint.

In the following study we took advantage of patients enrolled in the Swiss Spanish Intermittent Treatment Trial (SSITT). After successful ART for years, these patients underwent four cycles consisting of 2 weeks treatment interruption and 8 weeks retreatment before cessation of ART for a prolonged time period.

Here, we aimed at identifying shared genetic signature of rebounding viruses during STI in patients “controlling” viremia at levels below 5000 copies/ml of viral RNA. A strong evidence for the genetic signature 268E/358T associated with virologic control has been found. Next we tried to validate this finding, amongst others, in a group of well characterized primary HIV infected patients enrolled in the ZPHI-study. For that, I provided 1610 pre-treatment clonal C2-V3-C3 sequences of 104 ZPHI patients. I extracted viral RNA from plasma and did subsequent PCR amplification, molecular cloning and bidirectional sequencing of the C2-V3-C3 region of HIV-1 *env*. By using different bioinformatic tools, I edited and aligned sequences, in addition to manually adjusting. To exclude laboratory contamination I constructed Neighbor-joining phylogenetic trees by using HXB2 and Non-B strains as references and boostrapping. Patient’s characteristics were consolidated from different data sources.

The very strong positive predictive values of the signature 268E/358T associated with viral control in 9 SSITT patients were not confirmed in the 104 ZPHI patients analyzed during PHI. However, high negative predictive values were retained in this group.

Association between specific HIV-1 Env traits and virologic control *in vivo*

Beda Joos [a], Philip Rieder [a], Marek Fischer [a], Herbert Kuster [a], Peter Rusert [b], Alexandra Trkola [b], Satish K. Pillai [c], Joseph K. Wong [c], Rainer Weber [a], Huldrych F. Günthard [a]

[a] University Hospital Zürich, Division of Infectious Diseases and Hospital Epidemiology, Rämistrasse 100, CH-8091 Zürich, Switzerland

[b] Institute of Medical Virology, University of Zürich, Switzerland

[c] San Francisco Veterans Affairs Medical Center, University of California San Francisco, USA

Abstract

HIV RNA levels are influenced by genetic characteristics of both the host and the virus. Here we applied machine learning techniques to determine if plasma-derived HIV-1 amino acid sequences can be used to predict spontaneous virologic control. We studied the relationship between HIV-1 *env* genotype and viral load in 20 chronically infected patients undergoing treatment interruptions (SSITT, Swiss-Spanish Intermittent Treatment Trial) and in 104 primary HIV infected (PHI) patients before antiretroviral therapy (cART) and where applicable also after treatment stop. Extensive longitudinal sampling during the interruptions was performed in nine SSITT patients. Sequences obtained from these nine patients during the first virus rebound were used as a training data set and revealed a strong genetic signature (accuracy 98.6% in cross-validation) associated with control of viremia at levels below 5000 copies/mL of viral RNA maintained for at least 2 months after the final cART stop. The simple sequence pattern at gp120 positions 268E/358T was confirmed to be predictive of control in the clonal sequences originating from these patients during all subsequent rebounds. Sequences from the remaining 11 SSITT patients with less frequent sampling and from the PHI patients were used for external validation. High sensitivities (71–100%) and negative predictive values (80–100%) but low positive predictive values (12–40%) were achieved in the patient-wise analysis which was based on presence of the genetic pattern in all clones. These results suggest that presence of virus lacking the amino acid pattern 268E/358T is associated with VL >5000 at baseline of PHI and with low probability of spontaneous virologic control after treatment stop. Conversely, however, presence of 268E/358T does not predict control of viremia. These residues in HIV gp120 might affect *in vivo* HIV-1 fitness either at the level of Env function or influence susceptibility to adaptive or innate immune response.

Introduction

Empirical approaches to development of HIV vaccines have so far, despite huge efforts, not been successful. Among the major obstacles are the enormous genetic variability and the high degree of glycosylation of the outer HIV envelope proteins enabling the virus to escape from immune recognition. Moreover, there is a narrow window of opportunity to clear the initial infection because latency is established within the first weeks, and destruction of $CD4^+$ T cells begins early after infection (Johnston and Fauci, 2008). HIV induces a natural immune response involving neutralizing antibodies and cytotoxic T lymphocytes. However, these responses are typically too weak to eliminate the virus and the targeted epitope sequences are continuously mutated to evade the immune system. The envelope glycoproteins gp120 and gp41 comprise the key determinants presenting their surface to the host's humoral immune system. They inevitably undergo continuous evolution driven by positive selection, and are therefore the most variable gene products of HIV. Eventually mutations arise resulting in escape from host immune recognition (Yamamoto and Matano, 2008). These mutations often confer considerable fitness cost to the virus which may subsequently be alleviated by selection of compensatory mutations (Goulder and Watkins, 2004).

To explore intrapatient HIV evolution in the region of gp120 we conducted longitudinal clonal sequencing of the envelope C2-V3-C3 domain in HIV-1 infected patients with long-term suppression of plasma viremia who interrupted combination antiretroviral therapy (cART) within a well-controlled structured treatment interruption trial (Swiss Spanish Intermittent Treatment Trial). After successful cART for years, these patients underwent four cycles consisting of 2 weeks treatment interruption and 8 weeks retreatment before cART was completely discontinued for a prolonged time period. During the strictly defined short interruptions HIV RNA usually rebounded within 2 weeks (Fischer et al., 2003a). Patients who maintained plasma viremia at levels below 5000 copies/mL of viral RNA for at least 2 months after the final treatment stop were classified as ‘‘controllers’’ (Fagard et al., 2003). Analysis of the sequences derived from plasma before the initiation of antiretroviral therapy (cART) revealed that lower viral diversity was associated with spontaneous control of viremia after treatment stop as well as with reduced *in vitro* replication capacity and higher plasma titers of neutralizing antibodies against autologous virus (Joos et al., 2005).

Thus, viral properties appeared to influence disease progression being at least in part responsible for spontaneous control of viremia after cessation of cART. We also performed extensive longitudinal clonal studies of plasma HIV RNA to characterize the rebounding virus during the short interruptions and after the prolonged treatment break in a subset of patients (Joos et al., 2008). Phylogenetic analysis exploiting the temporal relation of rebounding virus to pre-treatment sequences was in accord with mono-or oligoclonal reactivation of distinct virus populations from long-lived latently infected cells in contrast to expansion of virus populations replicating at a low level. Furthermore, viral diversity was restored only very slowly after the final treatment stop. This implies that a sustained evolutionary bottleneck was induced by suppressive cART. In the long term, evolution of the envelope sequences spanning the V3 loop and its flanking regions appeared to be driven by positive selection (Joos et al., 2007). Interestingly the adaptive mutations occurred almost exclusively at solvent exposed amino acid positions on the virion surface which confer a high external accessibility according to the 3D structure of the envelope gp120 (Huang et al., 2005) suggesting that these sites evolved in response to selection pressure induced by neutralizing antibodies or adaptations in the receptor interaction necessitated by shifting cellular tropism. However, in this analysis we detected widely different patterns of positively selected sites within the C2-V3-C3 region of gp120 in individual patients by codon-based likelihood methods (Kosakovsky Pond and Frost, 2005).

Here we aimed at identifying common genetic characteristics of rebounding viruses across 20 chronically HIV-1 infected patients having undergone multiple planned treatment interruptions. We employed machine learning techniques to determine if a shared genetic signature could be found in plasma virus rebounding during the first interruption after years of continuous suppressive therapy. Next, we searched for specific signatures associated with high or low distance from the most recent common ancestor (MRCA) in the phylogeny which would reflect virus populations replicating at different speed. In addition, we sought to determine if plasma-derived HIV amino acid sequences could be used to predict spontaneous control of viremia. We found evidence of a link between HIV Env sequence and viral load by decision tree learning and aimed to validate this genetic signature by analyzing amino acid sequences originating from different patient groups.

2. Materials and methods

2.1. Study subjects

Blood samples were obtained from chronically infected patients enrolled into the Swiss-Spanish Intermittent Treatment Trial (SSITT) at the University Hospital Zürich. Two patients were excluded because they did not complete the study. We selected those patients from whom a pre-treatment plasma sample was available provided that PCR amplification was successful. All patients had long-term combination antiretroviral therapy (cART) and undetectable viral loads (<50 copies/mL) for at least 6 months before study entry and then underwent four consecutive STI cycles (2 weeks off and 8 weeks on treatment), followed by a longer fifth treatment interruption of indefinite duration. Patients were classified into ‘‘controlling’’ and ‘‘noncontrolling’’ groups. Control of viremia was defined as spontaneous maintenance of viral RNA <5000 copies/mL for at least 2 months during the first 4 months after the final cART stop (Trkola et al., 2003). Twenty patients were eligible for the current analysis, 9 of them were sampled intensively during the interruptions. The salient patient characteristics (Joos et al., 2005, 2007; Trkola et al., 2004) and results of the clinical trial (Fagard et al., 2003; Fischer et al., 2003a,b, 2004a,b; Metzner et al., 2003; Oxenius et al., 2002a,b) have been reported previously.

We also included 104 primary HIV infected (PHI) patients enrolled in the Zurich PHI Study (http://clinicaltrials.gov). All had well-documented acute or recent HIV infection presumably within the last 3 months based on the results of plasma HIV-1 RNA, HIV p24 antigen determination, Western blot, HIV-gp120 avidity assay and/or clinical symptoms of acute retroviral syndrome. The studied subjects include 92 men and 12 women, the vast majority infected with HIV subtype B (77%), CRF01_AE (10%), A (3%), and some with C, F1, G, CFR02_AG, and CFR12_BF. Plasma samples were analyzed as part of an ongoing study for all subjects at the time of enrolment before antiretroviral therapy was started and where applicable also later on when cART was stopped after a median of 18 (range 7–30) months in a subset of 28 patients.

Both studies were approved by the Ethics Committee of the University Hospital Zürich and written informed consent was obtained from all patients.

2.2. RNA extraction, RT-PCR, cloning and sequencing

Extraction of viral RNA from plasma and subsequent PCR amplification, molecular cloning and bidirectional sequencing of the C2-V3-C3 region of HIV-1 env was performed as described (Joos et al., 2005). By extracting HIV RNA from a total of 1 ml plasma our RT-PCR strategy was designed to maximize the probability that the sequenced clones are representative of the actual viral population present in plasma *in vivo*. The method was evaluated by testing the misincorporation rate during the RT step, the fidelity of the polymerase and reproducibility in analyzing identical patient samples (Joos et al., 2007). Representativity of the isolated viral quasispecies was confirmed by comparison with limiting dilution analysis. Similar quasispecies compositions were found in few samples tested in parallel by molecular cloning and by single genome analysis. This indicates that the used cloning method was not excessively affected by resampling bias (Joos et al., 2008). Approximately 16 individual clones were obtained at each time point. Overall 1753 clonal sequences were obtained from a total of 109 longitudinal samples across 20 SSITT patients (31–190 clones per patient) and 1610 pre-treatment clones were isolated from 104 PHI patients.

2.3. Phylogenetic inference

Sequences were edited and aligned using Lasergene software version 5.08 (DNASTAR Inc., Madison, WI) and manually adjusted in the correct reading frame to the following GenBank sequences: HIV-1 subtype B isolates HXB2 (accession number K03455), ADA (M60472), YU2 (M93258), BAL (M68893), JRFL (U63632), NL4-3 (U26942), subtype A isolate U455 (M62320), subtype C isolate ETH2220 (U46016), subtype D isolate ELI (K03454), subtype F isolate BR020 (AF005494), and the HIV group M envelope fragment ZR59a-2 (AF030527) isolated from a historical plasma sample collected in 1959 (Zhu et al., 1998). Neighbor-joining phylogenetic trees were inferred by MEGA version 4 (Tamura et al., 2007) using genetic distance estimates obtained from Tamura-Nei 6-parameter model. Bootstrap analysis was performed using 1000 replicates.

Maximum likelihood phylogenetic analyses were performed by DNAML on a parallel computing cluster using randomized input order, global rearrangements, and multiple

jumble options (PHYLIP Phylogeny Inference Package version 3.6 distributed by J. Felsenstein, Department of Genetics, University of Washington, Seattle). The Slatkin–Maddison test (Slatkin and Maddison, 1989), a cladistic measure of inter-population gene flow, was used to evaluate the degree of viral population shift between the different time points in the longitudinal analysis of clonal sequences from the 20 SSITT patients. Screening for intrapatient recombination was performed by GARD (Genetic Algorithm for Recombination Detection), using the single breakpoint detection method (Kosakovsky Pond et al., 2006). These analyses showed no evidence of recombination in the sequences from any individual in this study. No putative breakpoints could be identified, and a single phylogeny rather than multiple segment-specific phylogenies best described the data in all cases (AIC scores were not improved with multiple trees).

2.4. GenBank accession numbers

The reported sequences were deposited in GenBank under accession numbers EF424791–EF426096, DQ002058–DQ002345, AY656534–AY656549, AY656440–AY656471, AY656249–AY6563 12, AY375663–AY375678, AY375616–AY375630, and AY375568– AY375583.

2.5. Coreceptor usage prediction

Predictions of coreceptor use were made by the WetCat algorithm (Pillai et al., 2003) and by use of position-specific scoring matrices trained on the syncytium-inducing phenotype (PSSM SI/NSI) and on the X4 or R5 coreceptor phenotype (PSSM X4/ R5) (Jensen et al., 2003). The former method applies a support vector machine-based learning technique to classify the V3 loop amino acid sequences of the envelope glycoprotein according to sequences from viruses with known coreceptor phenotype. The latter derives a score by using precalculated matrices for the 35 amino acids forming the V3 loop. Both methods yielded comparable results. CXCR4-using virus was sporadically detected in the longitudinal sequences but in the combined analysis no clear switch in coreceptor usage was found.

2.6. Glycosylation sites

Potential N-linked glycosylation sites were assessed with the N-GlycoSite tool provided by Los Alamos HIV Sequence database (Zhang et al., 2004).

2.7. Positive selection detection

Identification of positively selected amino acid positions was performed by Single Likelihood Ancestor Counting, Fixed Effects Likelihood, and Random Effects Likelihood methods using the HyPhy software package (Kosakovsky Pond and Frost, 2005). Duplicate sequences were removed and the optimal timereversible substitution model was determined for each patient before the analysis.

2.8. Machine learning

Open source machine learning software Weka 3.5 (Waikato environment for knowledge analysis) was used to develop decision trees by the J48 algorithm (Frank et al., 2004). The minimum number of objects per leaf node was set at high values in all classification experiments to confine the complexity of the models and reduce overfitting bias (since total numbers of sequences varied the best fitting models with minNumObj values of 10–20 were selected). Prediction performance was evaluated by stratified 10-fold cross-validation method. The sensitivity (recall) is defined as TP/TP + FN, specificity (precision) by TP/TP + FP, where TP is the number of the true positives, FN the number of underpredictions and FP the number of overpredictions. The balanced F-score is the weighted harmonic mean of recall and precision given by the formula F =2 x (precision x recall)/(precision + recall). Kappa coefficients (observed coincidence -expected coincidence)/ ([1 - expected coincidence]) and the area under the ROC curve (receiver operating characteristic) were also used to measure the performance of the classification algorithms and are also reported.

2.9. Statistical analyses

For standard statistical analyses GraphPad Prism version 5 (GraphPad Software, San Diego, CA) was used. Nonparametric tests with Bonferoni adjustment for multiple comparisons were applied for group comparisons (Mann–Whitney) and correlations (Spearman). Two-sided Fisher's exact test was used to determine associations between categories. Positive predictive value (PPV) is defined as TP/TP + FP, negative predictive value (NPV) as TN/ FN + TN (TP = true positives; FP = false positives; TN = true negatives; FN = false negatives).

3. Results

To study the association between HIV-1 *env* genotype and viral load in chronically infected patients we took advantage of the Swiss-Spanish Intermittent Treatment Trial (SSITT) and followed 20 patients longitudinally. The first samples were obtained during the chronic phase of infection before any treatment was started. Subsequently patients received combination antiretroviral therapy for a median of 2.7 (range 1.5–3.4) years and then underwent 4 treatment interruption cycles (2 weeks interruption and 8 weeks retreatment) followed by cART stop. Clonal HIV-1 *env* sequences encompassing the C2-V3-C3 region of gp120 were isolated from pre-treatment plasma and from the last available sample following treatment stop. Nine of the 20 patients also had intense sampling during the interruptions. Overall, 1753 clonal sequences were derived from 109 samples collected over 6.5 (range 3.3–8.6) years (Joos et al., 2007, 2008). A Neighbor-joining phylogenetic tree including all sequences failed to show any evidence of cross contamination between study subjects or contamination with laboratory strains (Fig. 1). Phylogenetic trees were also constructed with the sequences of each patient alone by using much more computer intensive maximum likelihood method. This allowed us to infer the individual most recent common ancestral sequences (MRCA) and to explore the structure of the phylogenetic trees in regard to sampling time.

The relationship between sampling time and viral population diversity was determined in the nine patients analyzed before treatment as well as during the four structured treatment interruptions and sequentially thereafter (Fig. 2). Relatively homogenous virus populations emerged during the short 2-week interruptions (time 0–0.8 years) but at different time points often phylogenetically distinct virus populations appeared. After treatment cessation, diversity increased gradually with a substantial delay. On average the level of pre-treatment diversity was not achieved until 2.5 years after beginning of the study, when appreciable divergence from the virus populations observed before treatment was also apparent (Fig. 3). Evolution of the genetic distances to the MRCA in the nine frequently sampled patients is shown in Fig. 4. While the average branch lengths to the MRCA remained relatively stable over the observed time period, transient drops were observed during STI in all individuals. In comparison with the genetic distance to the MRCA at pre-treatment, marked reductions by>1% occurred in 7/9 patients (range 1.1–2.7%) suggesting

rebound of archived viral variants which were already present generations before the initiation of cART or even at the time of infection. Increasing branch lengths to the MRCA indicating ongoing evolution were observed in the majority of patients after the final treatment stop.

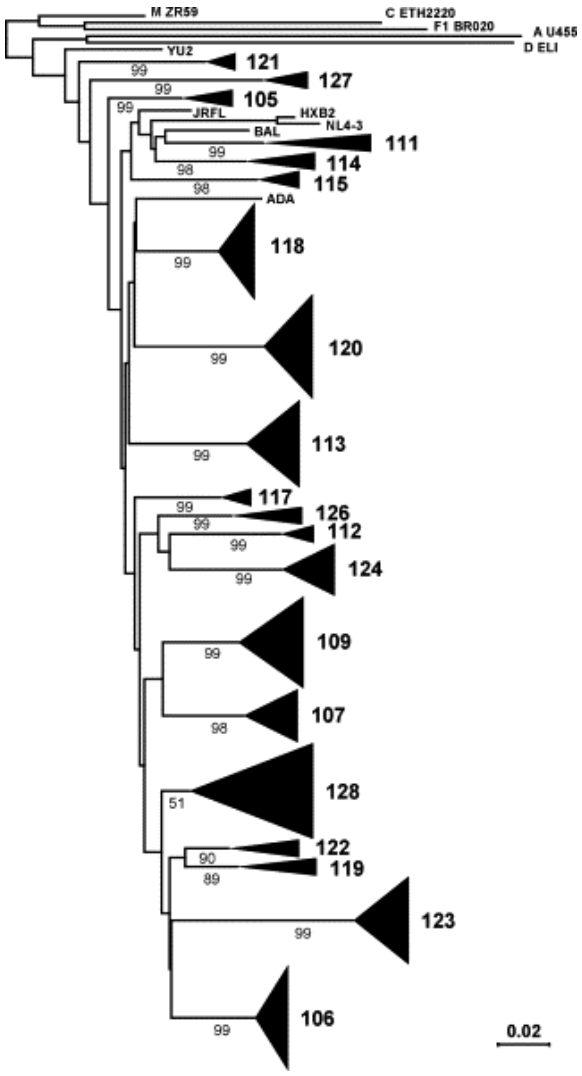

Fig. 1. Phylogenetic tree. Inferred neighbor-joining phylogenetic tree of clonal HIV-1 *env* C2-V3-C3 sequences derived from 20 patients. Elongated triangles represent the compressed subtrees containing 31–190 clones isolated from plasma of each patient before antiretroviral therapy and after treatment interruptions. The length of the triangle corresponds to the respective intrapatient diversity, thickness is proportional to the number of taxa. The bar denotes 2% nucleotide divergence and diversity (determined using the MEGA software), bootstrap values corresponding to 1000 replications are indicated below the branches.

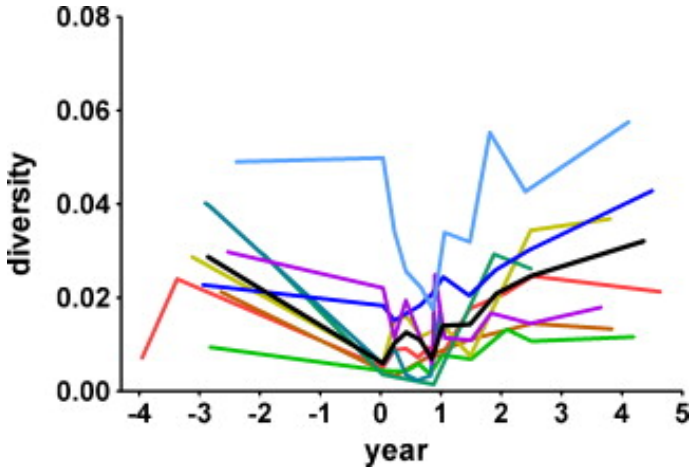

Fig. 2. Diversity. Longitudinal changes in viral diversity of clonal C2-V3-C3 sequences from nine patients before treatment and following structured treatment interruptions starting at time 0. The dark line indicates the median time course, patients have the same colour coding in this figure and Fig. 3 and Fig. 4. Distances were estimated by Tamura-Nei 6-parameter model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

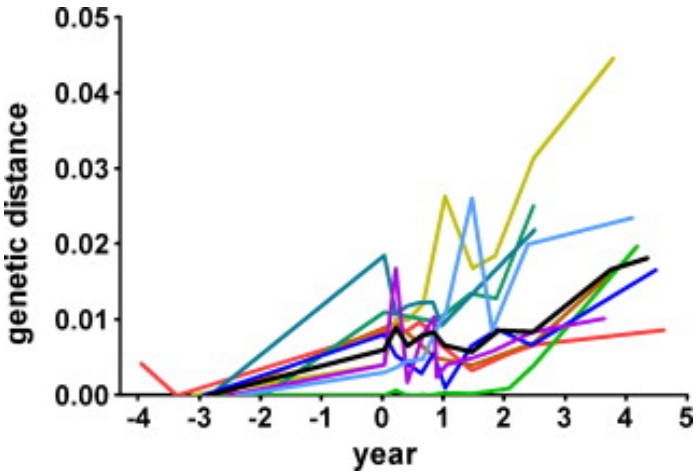

Fig. 3. Divergence. Longitudinal changes of average genetic distances (net divergence) from pre-treatment C2-V3-C3 sequences in nine patients following structured treatment interruptions starting at time 0. The dark line indicates the median time course, patients have the same colour coding in this figure and Fig. 2 and Fig. 4. Distances were estimated by Tamura-Nei model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

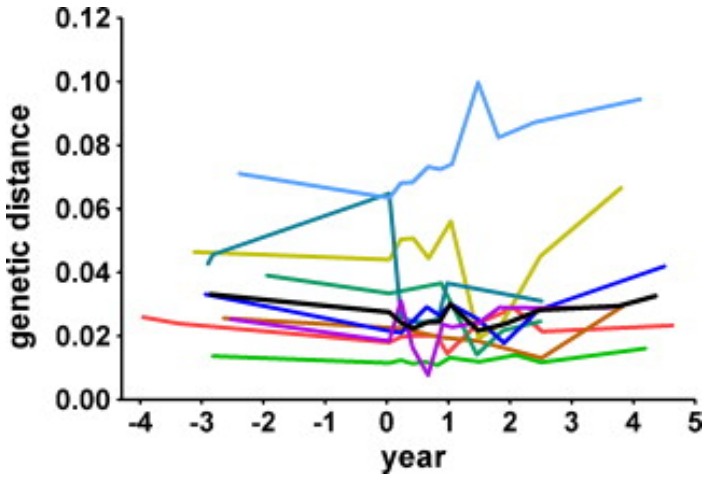

Fig. 4. Distance from MRCA. Longitudinal changes in average branch lengths to the most recent common ancestor (MRCA) of clonal C2-V3-C3 sequences from nine patients before treatment and following structured treatment interruptions starting at time 0. The dark line indicates the median time course, patients have the same colour coding in this figure and Fig. 2 and Fig. 3. Distances were estimated by Tamura-Nei model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

Table 1. Env C2-V3-C3 sequence classification based on control of viremia following treatment stop (cross-validation statistics).

Class variable	**control of viremia**		
Sequences	**first rebound**	**1st-5th rebound**	**all sequences**
Total number of instances	140	614	1753
Correctly classified instances	138 (98.6 %)	607 (98.9 %)	1743 (99.4 %)
Kappa coefficient	0.971	0.977	0.989
Sensitivity	1.000	0.993	0.996
Specificity	0.975	0.983	0.991
F- score	0.987	0.988	0.994
ROC Area	0.997	0.988	0.997

3.1. Genetic characteristics of rebounding virus

The distinct phylogenies and low diversities observed in plasma virus rebounding during the first treatment interruption after years of continuous suppressive therapy suggested potential reservoirs as the source of viral rebound. We posited that such virus might share some compartment-specific genetic signatures and features that could be obscured by interpatient sequence differences as has been shown for CSF virus and virus in genital secretions (Pillai et al., 2005, 2006). To explore this further, we employed machine learning techniques to search for a specific signature associated with first rebound among the clonal sequences originating from the nine patients with frequent longitudinal sampling.

Initially we compared the C2-V3-C3 sequences (126 amino acids) obtained during the first treatment interruption with those collected at pre-treatment, at later structured interruptions and during the last prolonged treatment interruption. In addition we also compared the first rebound sequences with those from all 20 patients at all time points. Testing showed no evidence of a genetic signature shared by the amino acid sequences derived from the first virus rebounds in plasma. Sensitivities of the predictions ranged from 0.357 to 0.736 and specificities from 0.71 to 0.893 resulting in balanced F-scores of 0.51–0.747. Kappa coefficients (chance-corrected proportional agreement) ranged from 0.488 to 0.641. Next, we applied decision tree learning to classify the sequences collected during the first STI on the basis of their distance to the MRCA. However, no clear genetic signature associated with high or low branch length to the MRCA could be detected.

Strong evidence for a genetic signature was found when the sequences derived from rebounding virus in the nine patients with frequent sampling were classified by control of viremia (defined by maintenance of viral loads <5000 RNA copies/mL for at least 2 months) after treatment stop (kappa coefficients >0.95, Table 1). Somewhat different classification models were inferred by using the sequences isolated during the first rebound alone (Fig. 5A) and those from the first to fifth rebounds (Fig. 5B) as training data sets. Interestingly, the common feature consisting of just glutamic acid at amino acid position 268 in combination with threonine at position 358 (268E/358T) was present in 647 out of 715 sequences obtained during all time points from controlling patients as opposed to only 63 of 687 sequences in non-controllers (Table 2, Fisher's exact $p < 0.0001$, kappa coefficient of agreement 0.813). The accuracy of

this association was consequently even slightly improved when the entire classification model (as illustrated in Fig. 5B) was linked to control of viremia in these patients (kappa 0.886; 670/715 sequences versus 35/687 sequences). A more complex, but also highly accurate decision tree was achieved when all envelope sequences from all individuals were tested. The signatures associated with control or no control of viremia primarily comprised the surface exposed amino acid residues at positions 268, 358, 269, 352 and 278 where significant positive selection has partially been detected before (relative surface accessibility 48.0 ± 18.8%, data shown elsewhere; Joos et al., 2007). Of note, this genetic signature found in patients spontaneously controlling plasma viremia was already present in the pre-treatment sequences and thus was not attributable to treatment interruptions.

3.2. Association of genetic signature with control of viremia

The depicted exploratory analysis of sequences isolated from the nine frequently sampled SSITT patients revealed that presence of the simple pattern 268E/358T in HIV envelope protein gp120 could be strongly associated with virologic control. Next we tried to validate this finding rigorously in different populations. For this purpose we evaluated clonal sequences from the remaining 11 SSITT patients with less frequent sampling and from a group of well-characterized primary HIV infected patients. These analyses were done by categorizing individual patients upon presence of 268E/358T in all clones isolated before treatment was started. We chose this restrictive condition under the assumption that a minority of fitter viruses might eventually overgrow the ‘‘beneficial’’ variants which presumably have reduced replication capacity.

While the predictive performance of the genetic pattern was naturally in the clonal analyses of the training data sets (Table 3,A– C), a different picture was found in the patient-wise analyses (Table 3, D–G). The very strong positive predictive values of the signature 268E/358T associated with viral control in the 9 frequently sampled SSITT patients were not confirmed in the 11 SSITT patients with less frequent sampling (Table 3, D). However, high negative predictive values were retained in this group as well as in the combined analysis of all chronically infected patients of the SSITT study (Table 3, E). Basically the same result, very high sensitivity and negative predictive value, was achieved by analyzing the sequences isolated from the first plasma samples of 104 primary HIV infected patients before treatment (Table

3, F). Participating in an ongoing study, most of the PHI patients currently remain under combined antiretroviral therapy. In 28 of them treatment was meanwhile stopped and viremia was contained at levels below 5000 copies/mL in some patients. Acceptable sensitivity and negative predictive value but low positive predictive value was also obtained in this group (Table 3, G). Interestingly, raising the cut-off for virologic failure from 5000 to 10,000 RNA copies/mL did practically not influence the result of the analysis. Five thousand was chosen here in order to be consistent with the definitions of the SSITT study.

The signature 268E/358T is present in 50% of the Env sequences which are collected in the Los Alamos National Laboratory HIV Sequence Database and represent the full spectrum of worldwide HIV-1 and SIVcpz sequences (year 2007 web alignment, 1766 sequences, (http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html). A fraction of 73% was found among the currently available sequences of subtype B viruses, in particular 71% and 79% of those with predicted CCR5 and CXCR4 coreceptor usage, respectively.

Table 2. Env C2-V3-C3 sequence classification according to control of viremia following treatment stop. Validation of patterns inferred from the training data sets (rebounding virus) in longitudinal samples collected from nine patients before treatment (pre), during 5 structured interruptions (STI), and 1-2 years as well as ≥2.5 years following treatment stop. Numbers indicate frequencies of matching clonal sequences.

Patient	Control of viremia	Time	268E & 358T		Signature associated with control	
			present	absent	present	absent
106	yes	pre	32	0	32	0
		STI	78	1	79	0
		1-2y	48	0	48	0
		≥2.5y	31	0	31	0
107	yes	pre	6	10	14	2
		STI	31	1	32	0
		1-2y	12	4	13	3
		≥2.5y	11	21	11	21
118	yes	pre	16	0	16	0
		STI	80	0	80	0
		1-2y	41	7	41	7
		≥2.5y	31	1	31	1
123	yes	pre	12	4	10	6
		STI	63	0	63	0
		1-2y	46	2	46	2
		≥2.5y	24	8	29	3
124	yes	pre	16	0	16	0
		STI	30	0	30	0
		1-2y	32	0	32	0
		≥2.5y	7	9	16	0
109	no	pre	10	22	10	22
		STI	1	94	1	94
		1-2y	21	2	21	2
		≥2.5y	0	16	0	16
113	no	pre	0	16	0	16
		STI	0	63	1	62
		1-2y	3	45	0	48
		≥2.5y	11	21	0	32
120	no	pre	0	16	0	16
		STI	0	94	0	94
		1-2y	0	47	0	47
		≥2.5y	0	32	0	32
128	no	pre	3	13	0	16
		STI	7	71	2	76
		1-2y	2	45	0	47
		≥2.5y	5	27	0	32
Total with control of viremia:			647	68	670	45
Total without control of viremia:			63	624	35	652

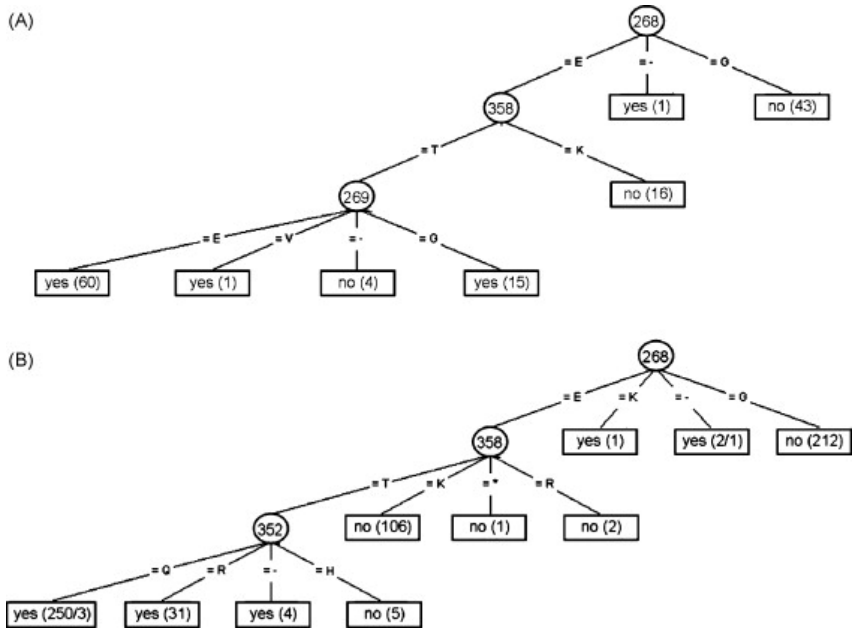

Fig. 5.Decision tree learning. Genetic characteristics of rebounding virus in nine patients with frequent longitudinal sampling. Signature associated with control of viremia after treatment stop (yes or no) in C2-V3-C3 sequences from first rebound (A, *n* = 140) and those from first to fifth rebound (B, *n* = 614). J48 decision tree learning algorithm and stratified 10-fold cross-validation by Weka. Residues are numbered according to HXB2 gp160 positions. Minus sign (−) denotes an amino acid deletion in the sequence. Numbers in parentheses indicate the total number of instances classified by the node and, if a second value is present, how many instances are incorrectly classified. Controlling patients are preferentially characterized by glutamic acid (E) at position 268 in combination with threonine (T) at 358.

Table 3: Association of genetic signature with control of viremia

Data set		pt (n)	CTR (n) [a]	clones (n)	type of analysis [b]	sensitivity	specificity	PPV [c]	NPV [c]
A	SSITT first rebound	9	5	140	cl	0.987	0.937	0.95	0.983
B	SSITT all rebounds	9	5	614	cl	0.993	0.976	0.972	0.994
C	SSITT all timepoints	9	5	1402	cl	0.905	0.908	0.911	0.902
D	SSITT external validation	11	2	176	pt	1	0.444	0.286	1
E	SSITT all patients	20	7	352	pt	0.714	0.615	0.5	0.8
F	PHI baseline	104	8	1610	pt	1	0.375	0.118	1
G	PHI plateau	28	8	429	pt	0.75	0.55	0.4	0.846

[a]) control of viremia <5000 copies/mL for at least 2 months after cART stop, resp. viral load <5000 (set F)

[b]) analysis based on individual clones (cl) or patients (pt)

[c]) PPV = positive predictive value, NPV = negative predictive value.

4. Discussion

It is well accepted that the replicative capacity of HIV in vivo is influenced by both genetic characteristics of the host and those of the virus. Interactions between virus and the host's immune system are complex and better understanding of the underlying mechanisms is important for planning optimized treatment strategies and for vaccine development. In a comprehensive longitudinal study we examined the virus characteristics of HIV envelope in a well-defined group of chronically HIV infected patients undergoing structured treatment interruptions. Patients were enrolled in the Swiss HIV-cohort study for years and were studied prior to cART, during continuous cART and subsequently during a structured treatment interruption trial (Swiss Spanish Intermittent Therapy Trial) (Fagard et al., 2003; Fischer et al., 2003a, 2004a; Joos et al., 2005, 2007; Oxenius et al., 2002a,b; Trkola et al., 2003). Here, we specifically attempted to find a common pattern in the phylogeny of the rebounding virus, to relate HIV-1 env genotype with HIV RNA levels and to test whether spontaneous virologic control after cessation of cART can be predicted from plasma-derived amino acid sequences.

Remarkable features could be deduced from the phylogenetic analyses (Joos et al., 2008). During the short treatment interruptions distinct mono-or oligoclonal virus rebounded but no significant evolution with regard to the MRCA was detected between the investigated time points. This indicates random reactivation of archived virus populations from latently infected cells without any significant residual replication in patients receiving potent cART. Unexpectedly, after the final treatment stop it took 1–2 years until the virus diversity reached the extent from before treatment. This long lasting effect shows that treatment introduced a strong bottleneck in viral evolution. Total divergence rate from the first to the last time points, averaged over the drug free periods only, was about 1% per year which is in agreement with previous reports (Shankarappa et al., 1999).

By applying machine learning techniques we further explored whether specific sequence patterns of HIV-1 env are associated with spontaneous virologic control. A decision tree inducer revealed in the sequences derived from the first virus rebound in plasma a strong genetic signature associated with maintenance of viremia at low levels after treatment stop. The signature was very simple and comprised two externally accessible amino acid sites located in the N-terminal and C-terminal

flanking regions adjacent to the V3 loop (Fig. 6). The pattern at gp120 positions 268E/358T was also found in the clonal sequences isolated during all subsequent rebounds. However, since all sequences originated from the same patients, this could also be due to overfitting. Indeed, attempts to verify the predictive capacity of the genetic pattern by external validation in a different group of SSITT patients and in primary HIV infected patients failed because specificity was lost. However, high sensitivity and negative predictive value were retained even in the large group of 104 PHI patients. This implies that presence of virus lacking the amino acid pattern 268E/358T was clearly associated with high viral loads at baseline and after treatment interruptions in both primary and chronically HIV infected patients. Conversely, however, presence of 268E/358T was not able to predict spontaneous control of viremia after treatment stop. It is conceivable that mutations at these two positions in HIV envelope gp120 affect in vivo fitness of the virus either at the level of Env function or by influencing susceptibility to adaptive or innate immune control.

A potential limitation of this study is that only a fragment of gp120 and not full length env sequences were analyzed. It is not clear whether the two identified positions alone contribute to differential replication capacity since the occurrence of compensatory mutations due to changes elsewhere in the envelope cannot be excluded. Additionally, the influence of host genetic polymorphisms and HLA types on the virologic outcomes could also not be assessed. Overall, the relatively small number of patients in this study may limit the power of our analysis and prohibit generalization.

The key finding of our analyses was that patients harbouring virus variants differing from gp120 268E/358T may have a lower probability of spontaneous virologic control after treatment stop. Intrinsic viral characteristics were, at least in part, responsible for reduced in vivo replication capacity. Whether incorporation of baseline genotypic analysis of HIV envelope will help in allocating different treatment options in the future, e.g. in patients who are not willing to continue lifelong therapy needs to be assessed in further trials.

Acknowledgements

The study was supported by the Swiss National Science Foundation (SNF) (grant no. 3345-65168 and 3100AO-103748/1 to HFG and grant no 320000-116035 to HFG), by the Swiss HIV Cohort Study (SNF grant no. 33CSC0-10878), SHCS project no. 290 and a grant from the EMDO Foundation (both to HFG), and a research grant from the Kanton of Zürich. The authors are grateful for the training received at the 14th International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology, September 2008, Cape Town. We thank the participating patients for their commitment, Christine Schneider, Christina Grube and Roland Hafner for excellent patient care, Esther Beerli, Friederike Burgener, Christine Leemann, Tuyet Trinh Lu, Barbara Niederöst and Bärbel Sauer for laboratory support and Ingrid Nievergelt and Christine Vögtli for administrative assistance.

References

Fagard, C., Oxenius, A., Gu¨ nthard, H., Garcia, F., Le Braz, M., Mestre, G., Battegay, M., Furrer, H., Vernazza, P., Bernasconi, E., et al., 2003. A prospective trial of structured treatment interruptions in human immunodeficiency virus infection. Arch Intern Med 163, 1220–1226.

Fischer, M., Hafner, R., Schneider, C., Trkola, A., Joos, B., Joller, H., Hirschel, B., Weber, R., Gu¨ nthard, H.F., 2003a. HIV RNA in plasma rebounds within days during structured treatment interruptions. Aids 17, 195–199.

Fischer, M.,Joos, B.,Hirschel, B.,Bleiber,G., Weber, R., Gunthard, H.F., 2004a. Cellular viral rebound after cessation of potent antiretroviral therapy predicted by levels of multiply spliced HIV-1 RNA encoding nef. J Infect Dis 190, 1979–1988.

Fischer, M., Joos, B., Wong, J.K., Ott, P., Opravil, M., Hirschel, B., Weber, R., Gunthard, H.F., 2004b. Attenuated and nonproductive viral transcription in the lymphatic tissue of HIV-1-infected patients receiving potent antiretroviral therapy. J Infect Dis 189, 273–285.

Fischer, M., Trkola, A., Joos, B., Hafner, R., Joller, H., Muesing, M.A., Kaufman, D.R., Berli, E., Hirschel, B., Weber, R., Gu¨ nthard, H.F., 2003b. Shifts in cell-associated HIV-1 RNA but not in episomal HIV-1 DNA correlate with new cycles of HIV-1 infection in vivo. Antivir Ther 8, 97–104.

Frank, E., Hall, M., Trigg, L., Holmes, G., Witten, I.H., 2004. Data mining in bioinformatics using Weka. Bioinformatics 20, 2479–2481. Goulder, P.J., Watkins, D.I., 2004. HIV and SIV CTL escape: implications for vaccine design. Nat Rev Immunol 4, 630–640.

Huang, C.C., Tang, M., Zhang, M.Y., Majeed, S., Montabana, E., Stanfield, R.L., Dimitrov, D.S., Korber, B., Sodroski, J., Wilson, I.A., et al., 2005. Structure of a V3-containing HIV-1 gp120 core. Science 310, 1025–1028.

Jensen, M.A., Li, F.S., van't Wout, A.B., Nickle, D.C., Shriner, D., He, H.X., McLaughlin, S., Shankarappa, R., Margolick, J.B., Mullins, J.I., 2003. Improved coreceptor

usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol 77, 13376–13388.

Johnston, M.I., Fauci, A.S., 2008. An HIV vaccine—challenges and prospects. N Engl J Med 359, 888–890.

Joos, B., Fischer, M., Kuster, H., Pillai, S.K., Wong, J.K., Boni, J., Hirschel, B., Weber, R., Trkola, A., Gu¨ nthard, H.F., 2008. HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proc Natl Acad Sci USA 105, 16725–16730.

Joos, B., Fischer, M., Schweizer, A., Kuster, H., Boni, J., Wong, J.K., Weber, R., Trkola, A., Gunthard, H.F., 2007. Positive in vivo selection of the HIV-1 envelope protein gp120 occurs at surface-exposed regions. J Infect Dis 196, 313–320.

Joos, B., Trkola, A., Fischer, M., Kuster, H., Rusert, P., Leemann, C., Boni, J., Oxenius, A., Price, D.A., Phillips, R.E., et al., 2005. Low human immunodeficiency virus envelope diversity correlates with low in vitro replication capacity and predicts spontaneous control of plasma viremia after treatment interruptions. J Virol 79, 9026–9037.

Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22, 1208– 1222.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D., 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22, 3096–3098.

Metzner, K.J., Bonhoeffer, S., Fischer, M., Karanicolas, R., Allers, K., Joos, B., Weber, R., Hirschel, B., Kostrikis, L.G., Gu¨ nthard, H.F., 2003. Emergence of minor populations of human immunodeficiency virus type 1 carrying the M184 V and L90 M mutations in subjects undergoing structured treatment interruptions. J Infect Dis 188, 1433–1443.

Oxenius, A., McLean, A.R., Fischer, M., Price, D.A., Dawson, S.J., Hafner, R., Schneider, C., Joller, H., Hirschel, B., Phillips, R.E., et al., 2002a. Human immunodeficiency virus-specific CD8(+) T-cell responses do not predict viral growth and clearance rates during structured intermittent antiretroviral therapy. J Virol 76, 10169– 10176.

Oxenius, A., Price, D.A., Gu¨ nthard, H.F., Dawson, S.J., Fagard, C., Perrin, L., Fischer, M., Weber, R., Plana, M., Garcia, F., et al., 2002b. Stimulation of HIV-specific cellular immunity by structured treatment interruption fails to enhance viral control in chronic HIV infection. Proc Natl Acad Sci USA 99, 13747–13752.

Pillai, S., Good, B., Richman, D., Corbeil, J., 2003. A new perspective on V3 phenotype prediction. AIDS Res Hum Retroviruses 19, 145–149.

Pillai, S.K., Good, B., Pond, S.K., Wong, J.K., Strain, M.C., Richman, D.D., Smith, D.M., 2005. Semen-specific genetic characteristics of human immunodeficiency virus type 1 env. J Virol 79, 1734–1742.

Pillai, S.K., Pond, S.L., Liu, Y., Good, B.M., Strain, M.C., Ellis, R.J., Letendre, S., Smith, D.M., Gu¨ nthard, H.F., Grant, I., et al., 2006. Genetic attributes of cerebrospinal fluid-derived HIV-1 env. Brain 129, 1872–1883.

Shankarappa, R., Margolick, J.B., Gange, S.J., Rodrigo, A.G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C.R., Learn, G.H., He, X., et al., 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol 73, 10489–10502.

Slatkin, M., Maddison, W.P., 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123, 603–613.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24, 1596–1599.

Trkola, A., Kuster, H., Leemann, C., Oxenius, A., Fagard, C., Furrer, H., Battegay, M., Vernazza, P., Bernasconi, E., Weber, R., et al., 2004. Humoral immunity to HIV-1: kinetics of antibody responses in chronic infection reflects capacity of immune system to improve viral set point. Blood 104, 1784–1792.

Trkola, A., Kuster, H., Leemann, C., Ruprecht, C., Joos, B., Telenti, A., Hirschel, B., Weber, R., Bonhoeffer, S., Gu¨ nthard, H.F., 2003. Human immunodeficiency virus type 1 fitness is a determining factor in viral rebound and set point in chronic infection. J Virol 77, 13146–13155.

Yamamoto, H., Matano, T., 2008. Anti-HIV adaptive immunity: determinants for viral persistence. Rev Med Virol 18, 293–303.

Zhang, M., Gaschen, B., Blay, W., Foley, B., Haigwood, N., Kuiken, C., Korber, B., 2004. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. Glycobiology 14, 1229–1246.

Zhu, T., Korber, B.T., Nahmias, A.J., Hooper, E., Sharp, P.M., Ho, D.D., 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 391, 594–597.

CHAPTER 5

Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection

Published as Kouyos *et al*., CID 2011; 1-8

Own contribution

An individual is typically infected by a few HIV-1 clones which dramatically increase in copy numbers but not in genetic diversity during the first month of infection. Following this acute infection period genetic diversity increase at a relatively linear rate with the length of infection. Thus, the diversity of the HIV-1 population within an infected patient may potentially classify an infection status into recent or chronic. This classification may be useful, for example, in epidemiological studies to evaluate the incidence of HIV-1. Here we evaluated the usefulness of ambiguous nucleotide calls from routinely generated HIV pol sequences as a potential measure of viral diversity. For this we investigated three different patient groups: patients with unknown date of infection (time of infection was calculated using a statistical backcalculation model, patients with known seroconversion and the ZPHI patients.

The study with the most stringent known date of infection was the the ZPHI data set. For each patient, a minimal and maximal timepoint of infection was estimated integrating all available information, including patient history relating to known risk situations, occurrence of first symptoms, previous negative test results, avidity assays and western blot.

Moreover, to evaluate the ZPHI set, the diversity of clonal *env* sequences of 130 patients with 10- 16 clones per sample have been used. I extracted viral RNA from plasma and did PCR amplification, molecular cloning and bidirectional sequencing of the C2-V3-C3 region of HIV-1 *env*. By using different bioinformatic tools, I edited and aligned sequences, in addition to manually adjusting. To exclude laboratory contamination I constructed Neighbor-joining phylogenetic trees by using HXB2 and Non-B strains as references and bootstrapping. The C2-V3-C3 derived clonal diversity was used to validate diversity as defined by ambiguous nucleotide calls from overall numbers of patients. Furthermore, it also allowed to test whether diversity measured on one gene (HIV pol) also correlated with the one of another gene (HIV env).

We observed a significant correlation between the fraction of ambiguous nucleotides in the *pol* gene and the diversity of clonal *env* sequences, which confirms that the fraction of ambiguous nucleotides is a good marker for viral diversity. In addition, the fraction of ambiguous nucleotides in the ZPHI data set is consistent with that found in the other data sets.

Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection

Roger D. Kouyos[1], Viktor von Wyl[1], Sabine Yerly[4], Jürg Böni[2], Philip Rieder[1], Beda Joos[1], Patrick Taffé[6], Cyril Shah[2], Philippe Bürgisser[7], Thomas Klimkait[8], Rainer Weber[1], Bernard Hirschel[5], Matthias Cavassini[7], Andri Rauch[10], Manuel Battegay[9], Pietro L. Vernazza[11], Enos Bernasconi[12], Bruno Ledergerber[1], Sebastian Bonhoeffer[3], Huldrych F. Günthard[1], and , the Swiss HIV Cohort Study

[1]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich

[2]Swiss National Center for Retroviruses, University of Zurich

[3]Institute of Integrative Biology, Eidgenössische Technische Hochschule Zurich, Zurich

[4]Laboratory of Virology, University Hospital Geneva and University of Geneva Medical School

[5]Division of Infectious Diseases, Geneva University Hospital, Geneva

[6]Swiss HIV Cohort Study Data Center

[7]Division of Immunology, University Hospital Lausanne, Lausanne

[8]Institute for Medical Microbiology, University of Basel

[9]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, Basel

[10]Clinic for Infectious Diseases, Bern University Hospital and University of Bern

[11]Division of Infectious Diseases, Cantonal Hospital St Gallen, St Gallen

[12]Division of Infectious Diseases, Regional Hospital Lugano, Lugano, Switzerland

Abstract

BACKGROUND: The time passed since the infection of a human immunodeficiency virus (HIV)-infected individual (the age of infection) is an important but often only poorly known quantity. We assessed whether the fraction of ambiguous nucleotides obtained from bulk sequencing as done for genotypic resistance testing can serve as a proxy of this parameter.

METHODS: We correlated the age of infection and the fraction of ambiguous nucleotides in partial *pol* sequences of HIV-1 sampled before initiation of antiretroviral therapy (ART). Three groups of Swiss HIV Cohort Study participants were analyzed, for whom the age of infection was estimated on the basis of Bayesian back calculation (n = 3,307), seroconversion (n = 366), or diagnoses of primary HIV infection (n = 130). In addition, we studied 124 patients for whom longitudinal genotypic resistance testing was performed while they were still ART-naïve.

RESULTS: We found that the fraction of ambiguous nucleotides increased with the age of infection with a rate of .2% per year within the first 8 years but thereafter with a decreasing rate. We show that this pattern is consistent with population-genetic models for realistic parameters. Finally, we show that, in this highly representative population, a fraction of ambiguous nucleotides of >.5% provides strong evidence against a recent infection event <1 year prior to sampling (negative predictive value, 98.7%).

CONCLUSIONS: These findings show that the fraction of ambiguous nucleotides is a useful marker for the age of infection.

Introduction

Human immunodeficiency virus type 1 (HIV-1) infections are initiated in most cases by a single virus [1], leading initially to a monomorphic viral population. Subsequently, viral diversity builds up gradually during HIV infection, first in a linear fashion but then at decreasing rates until a plateau is reached [2, 3]. In late-stage HIV infection, even decreases in viral diversity have been observed [3]. Thus, the diversity of the HIV population within an individual patient is potentially informative about the age of the infection, which is an important parameter because it allows an assessment of how far and how fast the infection has progressed. Seroconversion data are often lacking, and acute retroviral syndrome may have not occurred or may not have been recognized as such [4]. Thus, a method to estimate the infection duration based on viral sequences would be attractive, given the abundance of HIV sequence data from genotypic drug resistance tests.

Genotypic resistance tests use nucleotide sequences to infer to what degree different drugs may inhibit a given viral population of an HIV-infected individual. For economic reasons, these sequences are obtained by bulk sequencing; that is, the sequencing procedure is applied to a diverse sample of the HIV population. If the frequency of the most frequent nucleotide at a given position exceeds a threshold (typically around 80%), bulk sequencing returns the predominant nucleotide at this position. However, if this is not the case, then so-called ambiguous nucleotide calls are reported, implying that the patient harbors viral strains with different nucleotides at this locus. Thus, the fraction of ambiguous nucleotides is a measure of the degree of polymorphism of the HIV population within a patient, which in turn should scale with the age of infection. Here, we assess to what degree the proportion of ambiguous nucleotides correlates with the time elapsed between HIV infection and sampling for genotyping. To this end, we relate the fraction of ambiguous nucleotides to the age of infection derived with methods of different accuracy from 4 data sets.

Materials and Methods

We used previously analyzed nucleotide sequences from patients included in the Swiss HIV Cohort Study (SHCS) drug resistance database (see Kouyos et al [5] and GenBank accession numbers therein for a random sample of sequences). The SHCS is a nationwide, prospective, clinic-based cohort study with continuous enrollment and semiannual study visits [6, 7]. The SHCS has been approved by the ethical committees of all participating institutions, and written informed consent has been obtained from the participants. The SHCS drug resistance database contains the results of 13,201 genotypic resistance tests from 9,231 patients performed by the 4 laboratories engaged in HIV resistance testing in Switzerland, stored in a central database developed and hosted by SmartGene (Zug, Switzerland; Integrated Database Network System, version 3.5.0) [8]. Resistance data stem from routine clinical testing (60% of tests) and from tests performed retrospectively from frozen repository plasma samples (40% of tests). Retrospective sequencing was performed systematically by analyzing the earliest plasma sample available for each patient. All laboratories perform population-based sequencing of the full protease gene and at minimum codons 28–225 of the reverse transcriptase gene by means of commercial assays (Viroseq version 1, PE Biosystems; Virsoseq version 2, Abbott AG; vircoTYPE HIV-1 assay, Virco Lab) and in-house methods [9].

Subtype B dominates the Swiss HIV-1 epidemic [10, 11], and we therefore focused on this subtype (n = 9,157 sequences). Because antiretroviral therapy strongly distorts viral diversity [12], we included only sequences from patients who were therapy-naïve at the time of sampling and for whom an independent estimate of infection time was available (3,307 sequences, each from a different patient). The year of infection was estimated as described elsewhere [13] as the median of patient-specific infection time estimates based on a Bayesian back calculation model incorporating the dates of the first positive or last negative HIV test results and CD4 counts as predictor variables. Note that this infection time estimate is independent of the viral sequence. We refer to this data set as the full data set. Furthermore, we considered 3 additional sets of patients, for whom times or time differences are known with better accuracy: (1) In the set of seroconverters (366 patients), patients were recruited to the cohort within 1 year after infection, based on documented negative and positive HIV test results no longer than 180 d apart [11]; (2) The

longitudinal set contains 124 patients for whom sequences are available at 2 time points (at least 6 months apart). Although for this data set the time of infection is not more precise than for the large data set, the age difference of the 2 samples is known exactly; and (3) The Zurich Primary HIV Infection Study (ZPHI; ClinicalTrials.gov identifier, NCT00537966) set contains 130 patients with sequences obtained during acute infection (median time from infection, 41 d; interquartile range [IQR], 28–56 d) [14].

The relationship between the proportion of ambiguous positions (the dependent variable) and the time since infection was investigated using linear regression analysis. Because model residuals were not normally distributed, we performed a bootstrap analysis with 1,000 replicates to obtain 95% confidence intervals (CIs). We verified results by repeating all analyses on logit-transformed fractions by use of a generalized linear model for proportions. Both crude and adjusted analyses, using the patient's mode of HIV acquisition, ethnicity, sex, and age, were performed. In addition, a variable coding for the sequence-generating laboratory was included to account for assay- or laboratory-specific effects. All P values were 2-sided, and the level of statistical significance was set at .05.

We sought to identify a cutoff in the proportion of ambiguous sites to classify a patient's infection status into recent (infected for ≤1 year) or chronic (infected for >1 year). The classification performance of 2 categorizations of the proportion of ambiguous sites was evaluated with receiver operating characteristic (ROC) analyses. The first analysis included ambiguous sites as a categorical variable with 5 groups based on quintiles, whereas for the second analysis only 2 categories were included based on an a priori defined cutoff of ≤.5% or >.5% ambiguous positions.

Results

In this study, we included HIV subtype B sequences of 3,307 patients. As shown in Table 1, the majority of individuals in our sample were male (n = 2,593; 78.4%) and of white ethnicity (n = 2,933; 88.7%). Of these 3,307 patients, 1,482 (45%) had acquired HIV through homosexual contacts, 862 (26.1%) through intravenous drug use, and 857 (25.9%) through heterosexual intercourse. The median year of infection was 1995 with an IQR of 1991–2000, and the median duration from infection to sampling for genotypic sequencing was 4.7 years (IQR, 3.3–6.9 years).

We found that for HIV-1 sequences sampled before initiation of antiretroviral therapy, the fraction of ambiguous nucleotides increases significantly with the age of infection. This relationship is shown in Figure 1A for the full data set (3,307 patients) for whom the age of infection has been estimated with the back calculation algorithm (see Materials and Methods and Taffe and May [13]). Figure 1A shows that the fraction of ambiguous sites grows linearly with time in the first few years of infection and then flattens off. This effect can be described by including time since infection as a linear and a quadratic term in an adjusted regression model with the fraction of ambiguous nucleotides as outcome: if all patients are taken into account, both the positive linear and the negative quadratic components significantly improve the quality of the fit (model with linear term, r2 = .14; model with additional quadratic term, r2 = .17) (Figure 1B; Table 1). Because the initial increase in the proportion of ambiguous positions is almost linear during the first phase of infection, we further restricted our analyses to patients with an HIV sequence obtained in this phase of linear increase, that is, within the first 8 years of infection (n = 2,729). These adjusted analyses revealed a yearly increase in the proportion of ambiguous sites of ~.2 percentage points per year (Table 1), which was comparable with results from unadjusted regression analyses with yearly increases of .25% (95% CI, .23%–.28%) for the full sample and .19% (95% CI, .18%–.21%) for the sample restricted to patients with an infection duration of ≤8 years. Interestingly, Table 1 indicates that the mode of HIV acquisition significantly affects the level of diversity. In particular, intravenous drug users exhibit a larger fraction of ambiguous nucleotides than do both heterosexuals and men who have sex with men, which may suggest broader transmission bottlenecks for blood-transmitted than for sexually transmitted HIV.

Table 1

	N (%)	All patients	N (%)	Patients infected ≤8 years
N	3307 (100)		2729 (100)	
Time since infection (per year)		**0.195 [0.170 to 0.221]**		**0.158 [0.141 to 0.175]**
Time since infection (quadratic term)		**-0.007 [-0.008 to -0.005]**		NA
Female sex	714 (21.6)	0.034 [-0.053 to 0.121]	542 (19.9)	0.054 [-0.047 to 0.159]
Age by quartile; [min;max]				
[25 ; 30]	949 (28.7)	Referent	862 (31.6)	Referent
[33 ; 35]	789 (23.9)	**0.146 [0.060 to 0.233]**	605 (22.2)	**0.140 [0.047 to 0.233]**
[38 ; 41]	781 (23.6)	**0.232 [0.144 to 0.319]**	613 (22.5)	**0.247 [0.145 to 0.338]**
[45 ; 54]	788 (23.8)	**0.237 [0.146 to 0.328]**	649 (23.8)	**0.252 [0.153 to 0.348]**
Ethnicity				
White	2933 (88.7)	Referent	2455 (90.0)	Referent
Black	63 (1.9)	0.169 [-0.060 to 0.399]	58 (2.1)	0.080 [-0.122 to 0.290]
Hispano-American	97 (2.9)	**0.325 [0.141 to 0.509]**	89 (3.3)	**0.343 [0.140 to 0.546]**
Asian	57 (1.7)	0.090 [-0.147 to 0.328]	55 (2.0)	0.095 [-0.109 to 0.290]
Unknown ethnicity	157 (4.7)	-0.039 [-0.200 to 0.122]	72 (2.6)	0.001 [-0.186 to 0.211]
Mode of HIV acquisition				
Heterosexual risks	857 (25.9)	Referent	722 (26.5)	Referent
Intravenous drug use	862 (26.1)	**0.233 [0.142 to 0.323]**	595 (21.8)	**0.247 [0.140 to 0.357]**
Homosexual risks	1482 (44.8)	-0.066 [-0.154 to 0.021]	1326 (48.6)	-0.063 [-0.154 to 0.031]
Unknown risk	106 (3.2)	0.121 [-0.062 to 0.305]	86 (3.2)	0.186 [-0.006 to 0.416]
Laboratory				
Laboratory A	213 (6.4)	Referent	178 (6.5)	Referent
Laboratory B	1450 (43.8)	**0.407 [0.274 to 0.540]**	1176 (43.1)	**0.357 [0.246 to 0.456]**
Laboratory C	319 (9.6)	**0.404 [0.247 to 0.562]**	264 (9.7)	**0.312 [0.180 to 0.430]**
Laboratory D	1325 (40.1)	**0.787 [0.653 to 0.921]**	1111 (40.7)	**0.705 [0.594 to 0.803]**
Calendar year of sequencing (median [interquartile range])	2007 [2006-2008]	**0.119 [0.101 to 0.137]**	2007 [2006-2008]	**0.111 [0.093 to 0.127]**
Log10 HIV RNA copies/mL at time of GRT (by 33 percentiles) [min, max]				
[3.3 ; 4.0]	1013 (30.6)	Referent	829 (30.4)	Referent
[4.5 ; 4.8]	1013 (30.6)	**0.097 [0.017 to 0.176]**	859 (31.5)	0.043 [-0.050 to 0.128]
[5.2 ; 5.7]	1011 (30.6)	**0.136 [0.056 to 0.217]**	853 (31.3)	0.063 [-0.020 to 0.151]
HIV RNA missing	270 (8.2)	-0.106 [-0.235 to 0.023]	188 (6.9)	-0.130 [-0.273 to 0.026]
Constant term		-1.046 [-1.231 to -0.860]		-0.885 [-1.042 to -0.732]

1 unit = 1%

Table 1: Summary of the adjusted least squares regression for the full data set. The 95% confidence intervals were estimated with bootstrapping over 1000 replicates. Note that for the quadratic model (first column) both the linear and the quadratic term significantly improve the fit, but not when only the first 8 years of infection were considered Point estimates printed in bold-face were statistically significant by Wald test.

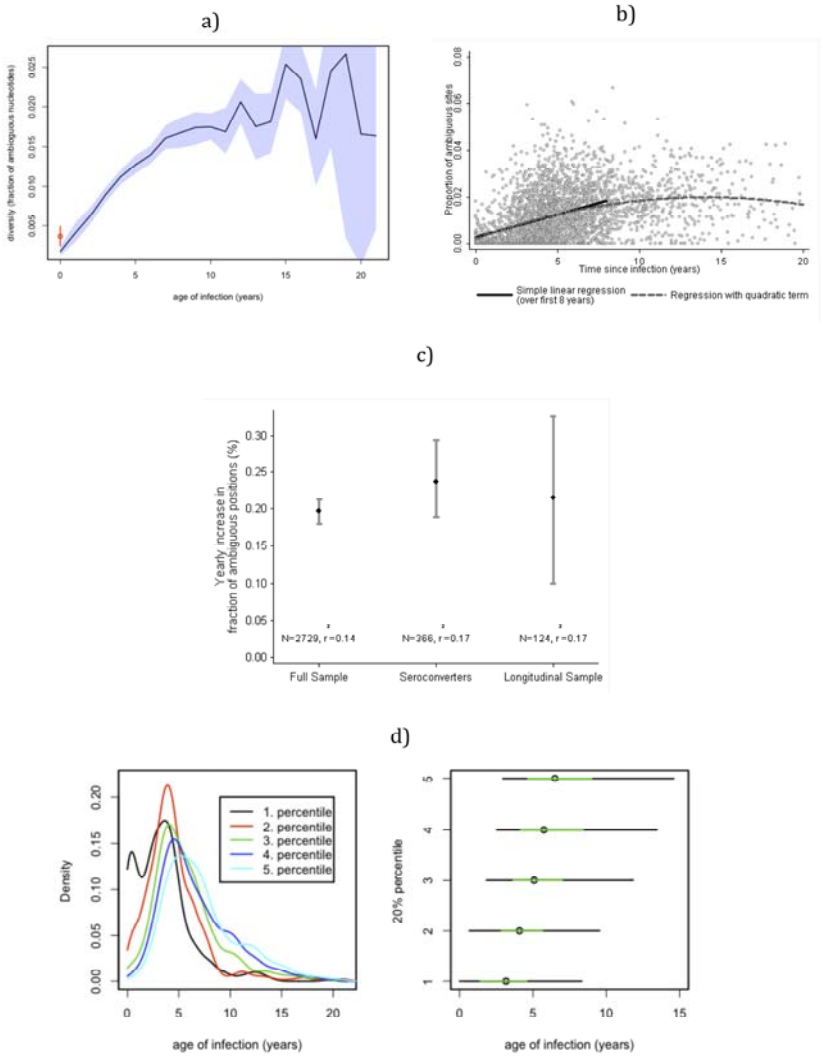

Figure 1.

Relationship between the year of infection and the fraction of ambiguous nucleotides (*f*). *A,* Mean of *f* as a function of the age of infection, where data points have been binned according to the age of infection in years (*n* = 3,307 patients). The shaded area corresponds to the 95% confidence interval of the means of *f*. These confidence intervals have been determined by bootstrap (with 1,000 samples). The red point corresponds to the mean of *f* for the Zurich Primary HIV Infection Study data set (*n* = 130), for which all sequences stem from the first few months after the infection. The associated red line gives the 95% confidence interval of this mean. *B,* Quadratic and linear fit of the full data set. Note that the linear fit is restricted to ages of infection of ≤8 years. *C,* Linear fit of the different data sets. Only sequences obtained within the first 8 years after infection were considered. *D,* Distribution of the age of infection (in years) for different fractions of ambiguous nucleotides. The left plot depicts the density plot of the age of infection for the 5 quintiles of *f*. The right plot depicts, for each of the 5 quintiles of *f,* the 25%–75% percentiles (*green lines*) and 5%–95% percentiles (*black lines*) of the year of infection.

To further assess the validity of the relationship between ambiguity and age of infection, we examined the smaller seroconverter set, which consists of 366 individuals who were recruited to the cohort in the early phase of their infection and for whom therefore the time point of infection was known within a given time interval of 6 months, and the longitudinal set, which contains 124 patients with sequence samples obtained from 2 time-points at least 6 months apart from each other. In the latter data set, inclusion was restricted to those patients for whom both samples were obtained within the first 8 years of infection. For these patients, we correlated the length of the time interval between the 2 measurements with the increase of the fraction of ambiguous nucleotides. Figure 1C shows that both additional data sets yield an estimate for the increase of ambiguous nucleotides over the first 8 years after infection that is similar to the estimate yielded by the original data set. Finally, we used the ZPHI data set, which contains the most stringent infection time estimate, to test the diversity around the time point of infection. Figure 1A shows that the fraction of ambiguous nucleotides in the ZPHI set is consistent with that found in the other 3 data sets. Moreover, for the ZPHI set, we observed a highly significant correlation (Spearman $\rho = 36$; $P < .001$) between the fraction of ambiguous nucleotides in the *pol* gene and the diversity of clonal env sequences [14], which confirms that the fraction of ambiguous nucleotides is a good marker for viral diversity.

In order to explore whether the observed correlation between infection time and the fraction of ambiguous sites could be exploited for age classification of individual HIV sequences, we subdivided our data set into quintiles of the fraction of ambiguous nucleotides f and inferred the distribution of the time of infection for each of these quintiles. Figure 1D shows that the fraction of ambiguous nucleotides provides useful information mainly on the lower bound of the age of infection: whereas sequences with $f < .68\%$ (ie, the first two quintiles) (Table 2) may stem from the early phase of an infection, this is unlikely for sequences with a larger fraction of ambiguous nucleotides. However, it is important to note that although it is unlikely that the HIV sequence from a recent HIV infection has a large fraction of ambiguous nucleotides, this may occur if the HIV population is founded by >1 virus, in which case diversity is large from the beginning on. This effect can be clearly seen in the ZPHI data set: a substantial minority of 24 (18%) of 130 patients exhibit a fraction of ambiguous nucleotides >.68%, even though all patients' samples have been sequenced in the

first months of their infection. This can be explained by results of Keele et al [1], who found that 23% of HIV infections are founded by >1 virus and hence should exhibit a large diversity even during primary infection.

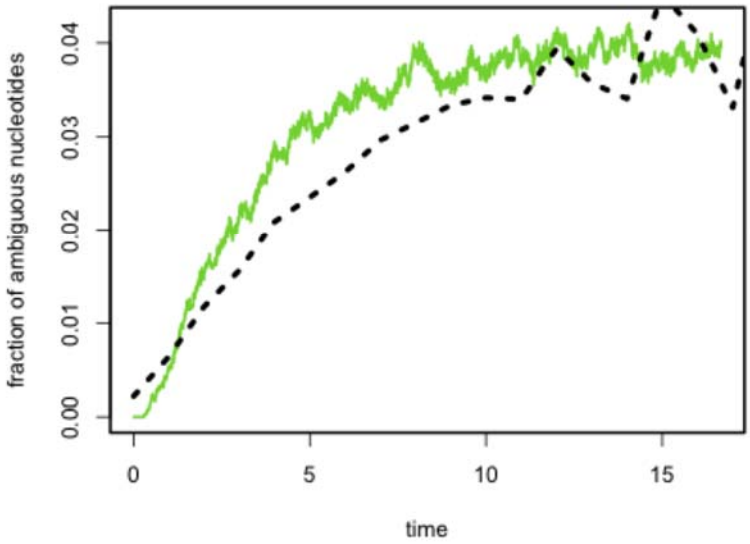

Figure 2.

Temporal increase of the fraction of ambiguous nucleotides in the Wright-Fisher model (WFM) for a population size of 500 and a mutation rate of 3×10^{-5} mutations per generation (*solid green line*) and at 4-fold degenerate third-codon positions in the full data set (*dashed black line*). The curve for the WFM has been obtained by averaging over 104 runs of the model. *N* and *m* denote the effective population size and the mutation rate, respectively. The WFM describes discrete and nonoverlapping generations in a population with fixed size *N*. Every generation, each of the *N* genomes undergoes mutation with probability *m*. Then the *N* genomes for the next generation are determined from the gene pool by drawing every offspring genome with uniform probability from the *N* parental genomes. Note that the WFM assumes selective neutrality.

Table 2

Strata number	Stratification cut-off	N per strata	N (%) of patients infected ≤ 1 year	Comparison of strata [1]	Sensitivity [2]	Specificity [3]	Correctly classified
	Pre-defined						
1	≤0.5%	1117	184 (16.5%)	1 vs. 2	86.8%	69.9%	70.9%
2	>0.5%	2190	28 (1.3%)				
	By quintiles						
1	0% to 0.15%	679	146 (21.5%)	1 vs. 2, 3, 4, 5	68.8%	82.8%	81.9%
2	0.16% to 0.67%	649	44 (6.8%)	1, 2 vs. 3, 4, 5	89.6%	63.2%	64.9%
3	0.68% to 1.35%	691	14 (2.0%)	1, 2, 3 vs. 4, 5	96.2%	41.4%	44.9%
4	1.36% to 2.07%	633	7 (1.1%)	1, 2, 3, 4 vs. 5	99.5%	21.1%	26.2%
5	2.08% to 6.65%	655	1 (0.2%)				

[1] Indicates how strata are collapsed for comparison, e.g. strata 1 vs. all remaining strata

[2] Sensitivity: Proportion of patients infected ≤1 year whose HIV sequence had ≤0.5% ambiguous positions

[3] Specificity: Proportion of patients infected >1 year whose HIV sequence had >0.5% ambiguous positions

Table 2: Summary of the two cut-off methods discussed in text: i) a priori defined threshold of 0.5% and ii) cutoffs based on quintiles.

We explicitly evaluated 2 different stratifications for cutoffs to predict whether a patient was infected for <1 year at the time of sampling for the genotypic test. As shown in Table 2, the selection of a cutoff of ≥.5% ambiguous positions predicted recent infections quite well in our full data set of 3,307 patients. Of 212 patients with a recent infection at the time of genotypic testing, 184 had ≤.5% ambiguous positions in their HIV sequence (sensitivity, 86.8%), but the specificity, which is the proportion of chronically infected patients with >.5% ambiguous positions, was only 70% (ROC area under curve, 78.3%). Therefore, this cutoff of ≤.5% may not be very useful to identify recently infected patients, but it can very accurately discriminate against patients with a chronic infection: only 28 (1.3%) of 2,190 patients with >.5% ambiguous positions had an infection duration of ≤1 year at the time of genotyping. Thus, if the observed fraction of ambiguous positions is >.5%, then there is a probability of 98.7% (95% CI 98.2%–99.1%) that the genotypic test was performed >1 year after infection (negative predictive value). Stratification by quintile did not improve the classification performance but essentially confirmed the above finding: when quintiles 1 and 2 are collapsed into a single category and compared with the remaining strata (ie, sequences with <.68% vs those with ≥.68% ambiguous positions), sensitivity increased to 89.6% but specificity decreased to 63.2%. When testing the .5% cutoff in the sample of observed seroconverters (n = 366), the performance was very similar to that in the full sample. Sensitivity and specificity were 84.3% and 59.8%, respectively (data not shown). Overall, 79.8% of patients were correctly classified, and the area under the ROC curve was 72.0%.

The observed increase in ambiguous positions over the time of infection likely reflects the diversification of the HIV population after the genetic bottleneck at the infection event. In the simplest scenario, this diversification is controlled only by mutation and genetic drift (ie, random extinction of viral strains). We modeled this scenario by simulating the Wright-Fisher model (WFM) [15], starting from an initially uniform population (ie, with no initial diversity). As the WFM assumes neutrality, we focused on ambiguous nucleotides at 4-fold degenerate third-codon positions for which the assumption of neutrality is best justified. We found that the WFM can reproduce the temporal increase of the fraction of ambiguous sites for parameters that have been shown to reproduce neutral evolution in HIV (see Kouyos et al [16] and references therein) (Figure 2). This finding implies that, at least for 4-fold degenerate third-codon positions, the increasing diversity of the HIV population can

be understood as the combined effect of mutation and genetic drift acting on a virus population, which is homogenous at the time point of infection.

Discussion

The most relevant clinical result of this study is that a large fraction of ambiguous nucleotides provides evidence against a recent infection event. In particular, we found that a threshold of .5% ambiguous nucleotides yields a negative predictive value of 98.7%. It is important to note that, by definition, the negative predictive value depends on the overall composition of the considered population at time of diagnosis. A strength of the sample studied here, however, is the high representativeness of the SHCS for the epidemic in Switzerland [6–8]. Thus, the composition of our sample is likely to be very similar to that of the population of HIV-1-infected individuals at the time of HIV infection diagnosis, and therefore the negative predictive value inferred here closely describes the situation in clinical practice. Finally, we find that although the frequency of ambiguous nucleotide calls differs significantly between laboratories (Table 1), the negative predictive value varies only marginally when calculated for samples from each laboratory separately (range, 98.5%–99.5%; χ^2 test, $P = .668$).

In order to test the robustness of our findings, we additionally assessed the impact of HLA types, transmitted resistance, and subtype. HLA types may affect viral diversity, for instance, by stabilizing selection (and possibly by maintaining detrimental viral mutations, as may be the case for type HLA-B*57) or, in the opposite, by leading to increased viral replication and diversity by triggering only suboptimal cytotoxic T lymphocytes responses, for instance, in the presence of homozygous HLA alleles. Using available HLA data from 352 individuals, we considered the homozygosity of HLA-B ($n = 32$) and the following HLA-B haplotypes: HLA-B*57 ($n = 23$), HLA-B*27 ($n = 25$), HLA-B*5801 ($n = 7$), and HLA-B*35 ($n = 80$). Upon inclusion of these variables in the adjusted regression model (restricted to a time of ≤8 years), neither of these HLA-B alleles reached statistical significance. There was a nonsignificant trend for higher viral diversity among carriers of homozygous HLA-B alleles of .243 (95% CI -.054 to .551). Concerning HIV subtype, it should be noted that the present analysis focused on subtype B, because most patients in the SHCS have been infected with that subtype. We found, however, very similar results for other subtypes (Table 3), indicating that our methodology extends beyond subtype B. However, it is also clear from the limited data available for non-B subtypes that additional tests are required for these subtypes. Finally, transmitted resistance mutations (present in 10% of the included patients) did not seem to have an impact when included in an

adjusted model analogous to that in Table 1 (with time restricted to ≤8 years), since the parameters were not statistically significant (data not shown). In summary, these 3 tests suggest that our findings are robust against HLA type, subtype, and transmitted resistance.

A potential limitation of our analysis is the accuracy of infection date estimates. We have confirmed the correlation between the age of an infection and the fraction of ambiguous base calls in several data sets with different estimation methods for infection duration. Moreover, we performed a sensitivity analysis to assess the impact of the uncertainty of infection date estimates from the back calculation model (data not shown). This was done by repeating the adjusted and unadjusted regression analyses on a random time point between the upper and the lower bound of the posterior infection date distribution [13]. Time trends were somewhat attenuated due to the introduction of additional variation with an increase of.13% per year (95% CI, .11%–.14%) for the unadjusted analysis and an increase of .11% per year (95% CI, .12%–.13%) for the adjusted analysis, but these point estimates still reached statistical significance.

We would like to emphasize that this study represents a proof of principle and that the details of the method should ideally be replicated and calibrated for each sequencing laboratory or method (especially in the light of our finding of significant differences between laboratories) (Table 1). Moreover, it is important to note that 10%–20% of recently infected patients do have a high viral diversity because they are infected with several strains. Therefore, the fraction of ambiguous nucleotides should be only one of several measurements used to decide whether a patient is recently infected.

Detection of ambiguous nucleotides is a byproduct of bulk sequencing. Here, we have shown that this byproduct carries important information on the age of the infection at sampling time. In particular, a large frequency of ambiguous nucleotides argues against a recent infection event. The qualitative pattern of an initially linearly increasing amount of diversity and subsequent saturation is consistent with the diversification pattern observed for the env gene [3]. Overall, our study highlights the usefulness of diversity measures as markers for infection age. This usefulness might further increase with the advent of array-based pyrosequencing, which allows the diversity of the HIV population to be analyzed in much more detail.

Acknowledgments

We thank the patients participating in the SHCS for their commitment, all the study nurses and study physicians for their invaluable work, the data center for data management, all the resistance testing laboratories for their high-quality work, and SmartGene for providing an impeccable database service. Furthermore, we thank Joseph Wong for critical reading of the manuscript.

References

1. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci USA 2008; 105:7552–7.
2. Bonhoeffer S, Holmes EC, Nowak MA. Causes of HIV diversity. Nature 1995; 376:125.
3. Shankarappa R, Margolick JB, Gange SJ, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol 1999; 73:10489–502.
4. Kahn JO, Walker BD. Acute human immunodeficiency virus type 1 infection. N Engl J Med 1998; 339:33–9.
5. Kouyos RD, von Wyl V, Yerly S, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. J Infect Dis 2010; 201:1488–97.
6. Schoeni-Affolter F, Ledergerber B, Rickenbach M, et al. Cohort profile: the Swiss HIV Cohort Study. Int J Epidemiol 2010; 39:1179–89.
7. Ledergerber B, Egger M, Opravil M, et al. Clinical progression and virological failure on highly active antiretroviral therapy in HIV-1 patients: a prospective cohort study. Lancet 1999; 353:863–8.
8. Von Wyl V, Yerly S, Boni J, et al. Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment–a comparison of different regimen types. Arch Intern Med 2007; 167:1782–90.
9. Yerly S, Vora S, Rizzardi P, et al. Acute HIV infection: impact on the spread of HIV and transmission of drug resistance. AIDS 2001; 15:2287–92.
10. Boni J, Pyra H, Gebhardt M, et al. High frequency of non-B subtypes in newly diagnosed HIV-1 infections in Switzerland. J Acquir Immune Defic Syndr 1999; 22:174–9.
11. Yerly S, von Wyl V, Ledergerber B, et al. Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey. AIDS 2007; 21:2223–9.
12. Joos B, Fischer M, Kuster H, et al. HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proc Natl Acad Sci U S A 2008; 105:16725–30.
13. Taffe P, May M. A joint back calculation model for the imputation of the date of HIV infection in a prevalent cohort. Stat Med 2008; 27:4835–53.
14. Rieder P, Joos B, von Wyl V, et al. HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men. AIDS; 24:1177–83.
15. Wright S. Evolution in Mendelian populations. Genetics 1931; 16:0097–159.
16. Kouyos RD, Althaus CL, Bonhoeffer S. Stochastic or deterministic: what is the effective population size of HIV-1? Trends Microbiol 2006; 14:507–11.

Synthesis

Despite an enormous body of knowledge accumulated in the last 30 years, a number of key questions regarding HIV-1 transmission, pathogenesis, and therapy remain unresolved. The primary HIV-1 infection (PHI) has attracted tremendous attention because of the importance of this stage of infection to all of these aspects of HIV-1. In addition it has been recognized that longitudinal studies of PHI patients may be key to better understanding of HIV-1 infection. Therefore, the Zurich Primary HIV-1 infection (ZPHI) – study is the basic for the present dissertation to address key questions of the early viral / host interaction. In addition to the PHI- patients the sources of the infection are of interest to understand transmission biology and viral and host contribution to the pathogenesis. Therefore, in this dissertation the identification of transmission pairs was one of the key aspects. We were able to link patients phylogenetically into transmission clusters and identified transmission events by using longitudinal clinical data. In contrast to the previous published literature we did a deep analysis of these transmission events with a surprising finding. A large number of new infections originated from patients in the chronic stage after stopping early ART. These findings argue for early, continuous ART not only for individual patient benefits but also for “treatment as prevention”.
These transmission clusters are the basic for further aims addressing mechanism and properties of viruses that are transmitted. In this project full length *env* clones from plasma HIV-RNA of transmitter-recipient pairs will be characterized. In detail, we will employ machine learning tools to look for evidence of a genetic signature by newly transmitted viruses. Then we will compare those specific codons with viral strains from transmitters and chronically infected patients and analyze if differences effectively exist between transmitted and non-transmitted viruses.

The second aim of the present dissertation is to define viral determinants explaining variations in viral setpoints reached in PHI patients after cessation of early ART. We studied viral diversity at the time of PHI by clonal sequencing of the HIV-1 envelope C2-V3-C3 region. We analyzed if differences in the mucosa are associated with differences in the complexity of the founder population. Specifically, we investigated if concomitant STI attenuated the genetic bottleneck and if single layered mucosal epithelium is more susceptible than multiple layered mucosas. In contrast to previous studies we did not found any differences in the complexity of the founder populations

between differences in the nature of the mucosal frontline. This indicates that other mechanisms than the mucosal barrier is involved in the genetic bottleneck.
Moreover, we wanted to test the hypothesis that higher viral diversity is associated with higher viral setpoints. However, because of the conclusions in project one that treatment is a prevention option and the improvements in medical treatment and drug tolerance, now, most patients stay on therapy. Therefore, too few viral setpoints reached after cessation of early ART could be analyzed for a significant conclusion about viral diversity during PHI as a surrogate marker for disease progression.

In the fourth chapter, we described a method for the optimal FH-probe design for qPCR in phylogenetic diverse organisms, such as HIV-1. All circulating and newly emerging variants of HIV-1 must be detected. In addition, FH-probes are very sensitive to mismatches which could lead to an underestimation of RNA copy numbers. We developed an algorithm to scan sequence databases for FH-probes in various genomic regions of HIV-1 approaching coverage of the global HIV-1 pandemic. Highly sensitive and accurate measurement of RNA transcripts is for example important for therapy monitoring and to evaluate the risk of HIV-1 transmission under ART as studied in chapter two.

Abbreviations

AIDS	acquired immunodeficiency syndrome
ARS	acute retroviral syndrom
ART	antiretroviral therapy
CA	capsid protein
cART	combination antiretroviral therapy
CCR5	chemokine receptor 5
CRF	circulating recombinant form
CTL	cytotoxict lymphocytes
CXCR4	CXC chemokine receptor 4
DNA	deoxyribonucleic acid
ELISA	enzyme-linked immunosorbent assay
env	envelope
FH-probes	fluorescent hydrolysis probes
gag	group-specific antigen
HAART	highly active antiretroviral therapy
HIV-1	human immunodeficiency virus type I
HSX	heterosexual
IN	integrase
IVDU	intravenous drug users
MA	matrix protein
MSF	men having sex with female
MSM	men having sex with men
MTCT	mother to child transmission
NC	nucleocapsid protein
nef	negative factor
NHP	non-human primate
NNRTI	non-nucleoside reverse transcriptase inhibitors
NRTI	nucleoside reverse transcriptase inhibitors
PHI	primary HIV-1 infection
PI	protease inhibitor
pol	polymerase
PR	protease

R5	CCR5 tropic viral strain
Rev	regulator of expression of virion proteins
RNA	ribonucleic acid
RT	reverse transcriptase
SHCS	Swiss HIV Cohort Study
SIV	simian immunodeficiency virus
STI	sexually transmitted disease
SU	surface unit
Tat	transcriptional transactivator
TM	transmembrane subunit
vif	virion infectivity
vpr	viral protein R
vpu	viral protein U
X4	CXCR4 tropic viral strain
ZPHI	Zurich Primary HIV-1 Infection

Acknowledgments

I am very grateful to my advisor at the USZ, Prof. Dr. med. Huldrych Günthard for letting me pursue my interests and for supporting me in any possible way during my studies and in my personal life.

Many thanks also to Prof. Alexandra Trkola and Prof. Andreas Wagner for the supervision and examination of my thesis.

I would also like to thank Beda Joos who always took time for me and my problems at work and was a great help. Further thanks go to Barbara Niederöst who supported me in the laboratory.

Other thanks go to all group members of the HIV-laboratory which was like a family: Claudia, Christine, Alexandra and Francesca for not only being colleagues but also friends, Noelia, for gossiping, Karin Metzner, always there for support, Herbert, Valentina, Wan Lin, Yannick.

I also thank our former laboratory members, especially, Sara Gianella.

Many people have contributed to the ZPHI-study and therefore to my doctoral thesis. I am very grateful to all patients participating in the ZPHI-study; Christina Grube, Barbara Hasse, Urs Karrer, Rolf Oberholzer, Elisabeth Presterl, Reto Laffer, Ulrich von Both, Klara Thierfelder, Dominique Braun, Yvonne Flammer, Markus Flepp and Thomas Frey for their dedicated patient care; Friederike Burgener and Dominique Klimpel for excellent laboratory assistance; Sara Gianella and Alexandra Scherrer for database support and data management.

Vielen Dank an meine Eltern, Walter und Elsa, und mein Bruder Dominic sowie für meinen besten Freund Rico und meine beste Freundin Claudia für die grossartige Unterstützung und, dass ihr über all die Jahre immer an mich geglaubt habt.

Curriculum Vitae

Full Name: **Philip Rieder**

Date of Birth: 09.11.1977

Nationality: Swiss

Citizenship: Bretzwil, BL

Education

Since 2007	PhD studies in virology and clinical trials
	MIM-PhD program, University and ETH Zurich, Switzerland
	Thesis: *Primary HIV-1 infection in Zurich: Molecular studies of transmission biology and epidemiology*
	Supervisor: Professor Dr. med. Huldrych F. Günthard, University Hospital Zurich
2005-2006	Master of Biology / Neuroscience, University of Zurich, Switzerland
	Thesis: *Characterisation of the cell adhesion molecule MDGA2A in the developing nervous system of Danio rerio*
	Supervisor: Dr. Matthias Gesemann, Brain Research Institute, University and ETH Zurich
2002-2005	Bachelor of Biology, University of Zurich, Switzerland
1999-2002	Matura, MAR, focus Mathematics and Physics
	Maturitätsschule für Erwachsene (AME), Aarau, Switzerland
1994-1998	Professional honor as „Vermessungszeichner“
	Bezirksgeometer, Zofingen

Honors and Awards

2010	Young Investigator Award
	17th Conference on Retroviruses and Opportunistic Infections, San Francisco
	Strong Effect of early Antiretroviral Therapy during Primary HIV-1 Infection in Preventing further Spread of HIV in Sexually Active Men Having Sex with Men
2008	Young Investigator Award
	15th Conference on Retroviruses and Opportunistic Infections, Boston
	Viral env Diversity and Transmission Clusters in Primary HIV-Infection: The Zurich Primary HIV Infection Study (ZPHI-study)

List of publications

Firs author or shared first author papers forming the base of this thesis

Rieder, P., B. Joos, V. von Wyl, H. Kuster, C. Grube, C. Leemann, J. Boni, S. Yerly, T. Klimkait, P. Burgisser, R. Weber, M. Fischer, and H. F. Gunthard. *HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men.* AIDS 2010, 24:1177-83.

Rieder P., Joos B, Scherrer U.A., Kuster H., Braun D., Grube C., Niederöst B., Leemann C., Gianella S., Metzner K.J., Böni J., Weber R., Günthard H.F. *Characterization of HIV-1 Diversity and Tropism in 145 Patients with Primary HIV-1 Infection***.** , Clinical Infectious Diseases, 2011, in revision

Althaus, C. F., S. Gianella, **P. Rieder**, V. von Wyl, R. D. Kouyos, B. Niederöst, A. Schmid, K. J. Metzner, B. Joos, H. F. Günthard, and M. Fischer. 2010. *Rational design of HIV-1 fluorescent hydrolysis probes considering phylogenetic variation and probe performance.* Journal of Virological Methods 165:151-160.

Co – author papers forming the base of this thesis

Joos, B., **P. Rieder**, M. Fischer, H. Kuster, P. Rusert, A. Trkola, S. K. Pillai, J. K. Wong, R. Weber, and H. F. Günthard. 2010. *Association between specific HIV-1 Env traits and virologic control in vivo.* Infection, Genetics and Evolution 10:365-372.

Kouyos, RD, von Wyl, V, Yerly, S, Boni, J, **Rieder, P**, Joos, B, Taffe, P, Shah, C, Burgisser, P, Klimkait, T, Weber, R, Hirschel, B, Cavassini, M, Rauch, A, Battegay, M, Vernazza, PL, Bernasconi, E, Ledergerber, B, Bonhoeffer, S, Gunthard, HF. 2011 *Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection.* Clinical Infectious Diseases. 2011 Feb 15;52(4):532-9.

Further co-author papers not mentioned in this thesis

Schmid, A., S. Gianella, V. von Wyl, K. J. Metzner, A. U. Scherrer, B. Niederöst, C. F. Althaus, **P. Rieder**, C. Grube, B. Joos, R. Weber, M. Fischer, and H. F. Günthard. 2010. *Profound Depletion of HIV-1 Transcription in Patients Initiating Antiretroviral Therapy during Acute Infection.* PLoS ONE 5:e13310

Stadler, T., R. Kouyos, V. von Wyl, S. Yerly, J. Böni, P Bürgisser, T. Klimkait, B. Joos, **P. Rieder**, W. Xie, HF. Günthard, A. Drummond, S. Bonhoeffer. 2010. *Estimating the basic reproductive number from viral sequence data.* submitted

Althaus C.F., Vongrad V., Niederöst B., Joos B., Di Giallonardo F., **Rieder P.**, Trkola A., Günthard H.F., Metzner K.J. & Fischer M.. *Novel targeted enrichment strategy enables detection of high variety of low abundance small noncoding RNAs in HIV-1 infection.* submitted